



# CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

## High-order correlations in species interactions lead to complex diversity-stability relationships for ecosystems

Elgin Korkmazhan and Alexander R. Dunn

Phys. Rev. E **105**, 014406 — Published 11 January 2022

DOI: [10.1103/PhysRevE.105.014406](https://doi.org/10.1103/PhysRevE.105.014406)

# High-order correlations in species interactions lead to complex diversity-stability relationships for ecosystems

Elgin Korkmazhan\* and Alexander R. Dunn

*Biophysics Program, Stanford University, Stanford, CA, USA and*

*Department of Chemical Engineering, Stanford University, Stanford, CA, USA*

(Dated: December 19, 2021)

How ecosystems maintain stability is an active area of research. Inspired by applications of random matrix theory in nuclear physics, May showed decades ago that in an ecosystem model with many randomly interacting species, increasing species diversity decreases the stability of the ecosystem. There have since been many additions to May's efforts, one being an improved understanding of the effect of mutualistic, competitive or predator-prey-like correlations between pairs of species. Here, we extend a random matrix technique developed in the context of spin-glass theory to study the effect of high-order correlations among species interactions. The resulting analytically solvable models include next-to-nearest neighbor correlations, that is two species changing their effect on each other due to other species they both interact with, as well as higher-order correlations. We find qualitative differences from May and others' models, including non-monotonic diversity-stability relationships. Furthermore, inclusion of particular next-to-nearest neighbor correlations in predator-prey as opposed to mutualist-competitive networks causes the former to transition to being more stable at higher species diversity. We discuss potential applicability of our results to microbiota engineering, and to the ecology of inter-predator interactions, such as cub predation between Lions and Hyenas as well as companionship between Humans and Dogs.

## INTRODUCTION AND BACKGROUND

Many ecosystems are composed of a large number of species that interact with each other in a plethora of ways. The extreme case of the Amazon rainforest is estimated to house several million species, ranging from carnivorous plants [1] to freshwater sharks [2]. Our own gut microbiome contains hundreds of microbial species that co-evolve with us [3–5]. The complexity of such ecosystems make it challenging to model and understand them [6].

A celebrated attempt to understand large ecosystems was made by May fifty years ago [7–9], which we detail here. Consider an ecosystem with  $N$  species, each with continuous, time-dependent abundance  $z_i(t)$  where  $i = 1, 2, \dots, N$ , which represent the elements of the abundance vector  $\vec{z}$ . The species interact with each other, which determines how their abundances change over time. Assume that the dynamics of  $z_i(t)$  are governed by a system of coupled ordinary differential equations admitting a steady state vector of species abundances  $\vec{z} = \vec{s}$ , around which the system can be linearized. Thus, the introduction of small changes in species abundances around the steady state, denoted  $x_i(t) = z_i(t) - s_i$  with corresponding vector  $\vec{x}$ , will dictate the dynamics according to

$$\frac{d\vec{x}}{dt} = \mathbf{M}\vec{x} \quad (1)$$

where  $\mathbf{M}$  is the Jacobian matrix evaluated at  $\vec{s}$ . Following precedent, we will refer to this matrix as the

*community matrix* [8] where element  $m_{ij}$  of  $\mathbf{M}$  encodes an *interaction* between the species pair  $j$  and  $i$ .

May considered a community matrix model of the form  $\mathbf{M}_1 = \mathbf{G} - \mathbf{R}$ . Here,  $-\mathbf{R}$  is a diagonal matrix representing the self-regulation of the species, which for convenience is taken to be  $-\mathbf{R} = -R\mathbf{I}$  where  $\mathbf{I}$  is the identity matrix and  $R$  is a positive constant.  $\mathbf{G}$  is a random matrix with independent and identically distributed entries: the entry distribution has mean 0 and finite standard deviation  $\sigma$  and is otherwise arbitrary. The parameter  $\sigma$  sets the range of the interaction strength between species pairs and we will simply refer to it as the *interaction strength* [10]. Note that  $\mathbf{M}_1$  is not symmetric; the effect of species  $i$  on  $j$  will differ from that of  $j$  on  $i$ . May focused on the *local stability* (hereafter referred to as *stability*) of the steady state species abundance of this ecosystem model. In general, small perturbations of abundances around a *stable* steady state are suppressed and the system returns back to the steady state, while those around an *unstable* one lead to divergence from the steady state. The stability can simply be determined from the eigenvalues of  $\mathbf{M}_1$ : if none of its eigenvalues have real part greater than 0, the steady state is stable. Thus, one can address the question of stability in May's model through the distribution of eigenvalues of  $\mathbf{M}_1$ .

According to the circular law conjecture [11, 12], in the limit of large  $N$ , the eigenvalues of  $\mathbf{G}$  are uniformly distributed on the disc of radius  $\sigma\sqrt{N}$  centered at the origin in the complex plane, with the largest real part of its eigenvalues consequently being  $\sigma\sqrt{N}$ . The circular law conjecture has been proven in its most general form only a decade ago by Tao, Vu and Krishnapur [13], who provide a universal result agnostic to the details of the matrix entry distribution. Thus, a simple change of variables in the eigenvalue distribution for  $\mathbf{G}$  leads

---

\* elgink@stanford.edu

to  $\sigma\sqrt{N} - R$  for the largest real eigenvalue part of  $\mathbf{M}_1$ . This implies that the ecosystem is stable when  $\sigma\sqrt{N} < R$ . One way to incorporate the fact that not all species pairs necessarily interact with each other is to introduce a connectivity fraction  $C$ , which randomly sets a fraction  $C$  of the entries of  $\mathbf{G}$  to 0. This simply modifies the inequality to  $\sigma\sqrt{NC} < R$ , bringing us to the particular *diversity-stability relationship* proposed by May [7], which emphasizes a trade-off among  $\sigma$ ,  $N$  and  $C$  within this ecosystem model. Thinking of the species interaction strength  $\sigma$  as a control parameter, we can define a *critical interaction strength*  $\sigma_c$  after which the ecosystem is always unstable [10]. For May's model,  $\sigma_c = \frac{R}{\sqrt{NC}}$  (see Fig. 1a where connectivity  $C$  is dropped following rest of the manuscript—see Results). It is important to note that this analysis does not say anything about "feasibility", i.e. having physically allowable (positive) values for the abundances in the steady state vector around which the system is linearized. The study of feasibility requires consideration of the original system before linearization, which the models in this manuscript will be mostly agnostic to. Relatedly, the models discussed in this manuscript, including those by May [7] and ref. [14], do not take into account any effect of the steady state vector of abundances on the Jacobian matrix at steady state, and hence any effect of it on stability [8]. It is interesting to contrast this to the case of simple generalized Lotka-Volterra models, where species interactions are typically linear with respect to species abundances [8, 15] and the steady state Jacobian matrix accordingly has an explicit dependence on the steady state abundances. We refer interested readers to a fraction of many revealing pieces of work on different models regarding feasibility and stability by refs. [16–20].

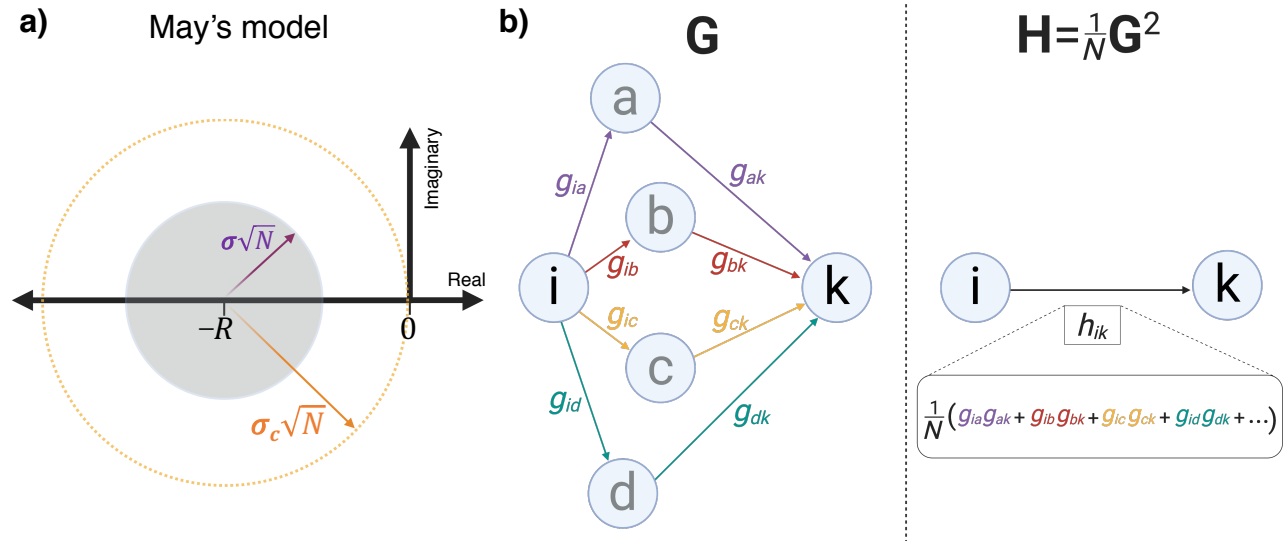
While the assumptions underlying May's approach have been questioned [8, 9, 17, 21–23], it provides an interesting null model and has inspired further investigations using random matrix approaches. An avenue of interest in ecology is the study of the effects of correlations among species interactions, which is assumed to not exist in May's model. This is potentially one way nature avoids May's diversity-stability constraint, with species interactions in different ecosystems having empirically different correlations [21, 24–26]. A common example of correlations in an ecosystem can be seen in ecosystems rich in predator-prey (also known as exploitive interaction) pairs, where it is expected to find more pairs of species that directly affect each other in an opposite manner, that is, through an interaction beneficial for one but detrimental to the other. Beyond that, it is possible that there are prevalent higher-order correlations in the direct interactions of species, such as those between the common predators of a prey, though this is not well-understood at the ecosystem level, potentially due to difficulties in experimental ecology (see Discussion).

An important case of correlations in ecosystems was addressed by Allesina and Tang [14] in a May-like model

where diagonally symmetric elements of the community matrix were set to have pairwise correlations, by drawing the pair of elements from a general bivariate distribution with correlation  $\rho$ . Through theory and simulations, they showed that the functional form of May's diversity-stability relationship undergoes simple modifications in the presence of such correlations. For example, a perfect predator-prey limit of the community matrix ( $\rho$  chosen such that  $m_{ij}m_{ji} < 0$  with marginal means of interactions 0) was shown to increase the critical interaction strength  $\sigma_c$  to  $\frac{R}{\sqrt{NC(1-|\rho|)}}$ , compared to the  $\frac{R}{\sqrt{NC}}$  in May's model. Empirical evidence suggests that correlations contribute to stability in more ways than described in ref. [14], but the theoretical basis is not clear [24]. There have indeed been other computational and theoretical advances on the effect of correlations among species interactions [26–31]. However, the topic is far from exhausted and enjoys only a small collection of analytical results [21]. More generally, the addition of complexity to models of large numbers of interacting species has proven challenging, but is critical to understand and make predictions regarding complex ecosystems [3, 5, 6].

While random matrix theory has helped achieve milestones in the modelling of ecological systems [9, 10, 14, 17, 20, 26], it has also been extensively applied in finance [32], neural information processing [33] and throughout physics [34–39]. One application includes the introduction of a "squared interaction matrix" to study the effect of next-to-nearest neighbor correlations in a fully connected spin-glass model in a simple, analytically tractable manner [40]. Here, we extend this approach to study 3-species (next-to-nearest neighbor) and higher order correlations in direct inter-species interactions using a particular set of correlation structures, which are easily generated by incorporating random matrices raised to integer powers in our models. For clarity, we would like to note that we do not study indirect interactions [41–43]. Indirect interactions arise, for example, when species A directly interacts with species B and alters the abundance of species B, which then leads to species A indirectly affecting the other species C that is directly interacting with species B ("linked chains of direct interactions") [42]. We also note that the high-order correlations we study are distinct from what are defined as high-order "interactions", where the direct interaction between two species is affected by the presence of a third species [10, 44–46] (this at times has been referred to as an indirect interaction as well [42]; for a rigorous discussion of high-order interactions, see ref. [47]). In contrast, we instead study a system of linearized differential equations, with high-order correlations among terms corresponding to the direct pairwise interactions of species. We show that with one correlation structure, the addition of high-order correlations beyond 3-species to May's model (such as *the friend of the enemy of my enemy helps me*)

FIG. 1. May's model and 3-species correlations



(a) The eigenvalues of May's community matrix  $\mathbf{M}_1$  are uniformly distributed in a disc of radius  $\sigma\sqrt{N}$  centered at the point  $(-R, 0)$  on the real axis in the complex plane. We have dropped the connectivity fraction  $C$  in May's original model for the remainder of the manuscript. As drawn, the system is stable. When  $\sigma > \sigma_c$  (purple radius larger than orange radius), the largest real eigenvalue component is larger than 0, and the ecosystem is unstable. (b) The inter-species interactions encoded in  $\mathbf{G}$  can be represented in a weighted directed graph (left panel) where each node represents a species and each directed link weight equals the interaction effect of one species on another. Here, only one of the two links (one direction) between species pairs is shown. Left panel shows an arbitrary subset of the 3-species paths from species  $i$  to  $k$ . In the right panel,  $h_{ik}$  is the modification to the interaction effect of  $k$  on  $i$  introduced when going from May's model  $\mathbf{M}_1$  to  $\mathbf{M}_{2,1}$ .

can cause the critical interaction strength to vary non-monotonically with species diversity. Thus the maximum critical interaction strength is achieved at finite values of species diversity, unlike in May's model. We then show that for a related correlation structure, predator-prey networks transition to being more stable than mutualist-competitive networks at higher species diversity. Overall, the correlation structures in our linear models give rise to analytically tractable, complex diversity-stability relationships. We discuss how our results may fit into broader ecological and evolutionary frameworks.

## RESULTS

We build on May's model [7], where we drop the connectivity fraction  $C$ , which gives the diversity-stability relationship  $\sigma\sqrt{N} < R$ . Following literature, we refer to  $\sigma$  as the interaction strength and focus on the critical interaction strength  $\sigma_c$  above which the ecosystem is always unstable for given  $N$  [10]. In May's model,  $\sigma_c = \frac{R}{\sqrt{N}}$  where the critical interaction strength

is seen to decrease linearly with increasing  $\sqrt{N}$  (Fig. 1a). Throughout the manuscript, we analyze modifications to May's model, and keep working in the large  $N$  limit. Finally, we refer to entries of  $\mathbf{G}$  in May's model as pairwise *interactions* between species unless we need to distinguish  $g_{ij}$  from  $g_{ji}$ . In this case, we refer to  $g_{ij}$  as the *interaction effect* of species  $j$  on  $i$ .

### Section I: positive n-species correlations

Now, instead of May's model  $\mathbf{M}_1 = -\mathbf{R} + \mathbf{G}$ , we first would like to analyze a modified ecosystem with community matrix

$$\mathbf{M}_{2,\beta} = -\mathbf{R} + \mathbf{G} + \frac{\beta}{N} \mathbf{G}^2 \quad (2)$$

where again, each matrix is of dimensions  $N \times N$  with  $N$  being species diversity.  $\beta$  is a positive constant ( $0 < \beta \leq 1$ ) determining the level of mixing of the rightmost term with the rest of the terms. We will be looking at the effect of increasing  $\beta$ , which we will show

to correspond to increasing next-to-nearest neighbor correlations in  $\mathbf{M}_{2,\beta}$ . Note that for a given instantiation of the random community matrix  $\mathbf{M}_{2,\beta}$ ,  $\mathbf{G}^2$  is simply the square of the random matrix  $\mathbf{G}$  whose entries are already drawn. First, let us focus on the rightmost term,  $\frac{\beta}{N}\mathbf{G}^2$ . The introduction of  $+\frac{\beta}{N}\mathbf{G}^2$  implies a particular next-to-nearest neighbor correlation structure in the ecosystem. Each element  $h_{ik}$  of the matrix  $\frac{1}{N}\mathbf{G}^2$  by definition equals to  $\frac{1}{N}\sum_j g_{ij}g_{jk}$  where  $g_{ij}g_{jk}$  is the product of the interaction effect of species  $j$  on  $i$  (in  $\mathbf{G}$ ) and the interaction effect of species  $k$  on  $j$  (in  $\mathbf{G}$ ). Thus the element  $h_{ik}$  of  $\frac{1}{N}\mathbf{G}^2$  represents the average of all pairwise interaction effects (in  $\mathbf{G}$ ) of  $k$  on other species, each weighted by the interaction effect (in  $\mathbf{G}$ ) of those species on species  $i$ .  $N$  corresponds to the number of 3-species *paths* between the two species  $i$  and  $k$  (Fig. 1b). Finally, note that the operator norm of  $\frac{1}{N}\mathbf{G}^2$  is  $\frac{\sigma}{\sqrt{N}}$  fold that of  $\mathbf{G}$ . Dean and Ritort, in a similar vein, introduced  $\mathbf{G}\mathbf{G}^T$  (instead of  $\mathbf{G}^2$ ) to the respective Hamiltonian in the squared interaction matrix Sherrington-Kirkpatrick model of spin-glasses to study another form of next-to-nearest neighbor correlations [40].

Coming back to the community matrix  $\mathbf{M}_{2,\beta} = -\mathbf{R} + \mathbf{G} + \frac{\beta}{N}\mathbf{G}^2$ , we see that increasing  $\beta$  decreases the critical interaction strength  $\sigma_c$  at which the ecosystem destabilizes (Fig. 2a). This is not surprising since we are simply increasing the largest real eigenvalue component of May's community matrix by  $\beta\sigma^2$ . This is seen from a fact that we will be repeatedly using throughout the manuscript: in general, if  $v$  is an eigenvalue of  $\mathbf{G}$ , and  $\chi$  is a polynomial,  $\chi(v)$  is an eigenvalue of  $\chi(\mathbf{G})$  (see Appendix). In this case,  $\sigma_c = \frac{1}{2\beta}(\sqrt{4\beta R + N} - \sqrt{N})$  which for large  $N$  behaves like the critical interaction strength of May's model,  $\frac{R}{\sqrt{N}}$ . Now, let us set  $\beta = 1$ ,  $\mathbf{M}_2 \equiv \mathbf{M}_{2,1}$  and consider the inclusion of higher-order correlations in May's model, that is longer species *paths* when modifying the direct inter-species interactions. Instead of only including 3-species correlations, we will include up to  $(F+1)$ -species correlations with the analogous structure, which gives the community matrix

$$\mathbf{M}_F = -\mathbf{R} + \mathbf{G} + \sum_{p=2}^F \frac{1}{N^{p-1}} \mathbf{G}^p \quad (3)$$

. The term  $\mathbf{G}^p$  takes into account all possible  $(p+1)$ -species paths between a pair of species, which total to  $N^{p-1}$  terms. In the case of  $\mathbf{M}_3$  we find that  $\sigma_c$  no longer monotonically decreases with  $N$ , as opposed to the case of  $\mathbf{M}_2$ , and it peaks at a non-zero species diversity (Fig. 2b; see Appendix). This is interesting in part due to its qualitative difference from May's model where the highest possible critical interaction strength occurs at species diversity  $N = 0$ .

We should note that the observed non-monotonicity in  $\sigma_c$  in the case of  $F = 3$  is, as expected, due to

the emergence of a high integer power of  $\sigma$ , with a cooperating factor of  $N$ , in the equations to solve for the largest eigenvalue real component and thus is not particularly special to the inclusion of correlations (see Appendix, Section I details). However, it is possible that the correlation structure we use is a natural way in which those terms can arise (see Discussion). In any case, we see that this approach allows for relatively easy analytical treatments of random matrix models with high-order correlations among the direct inter-species interactions.

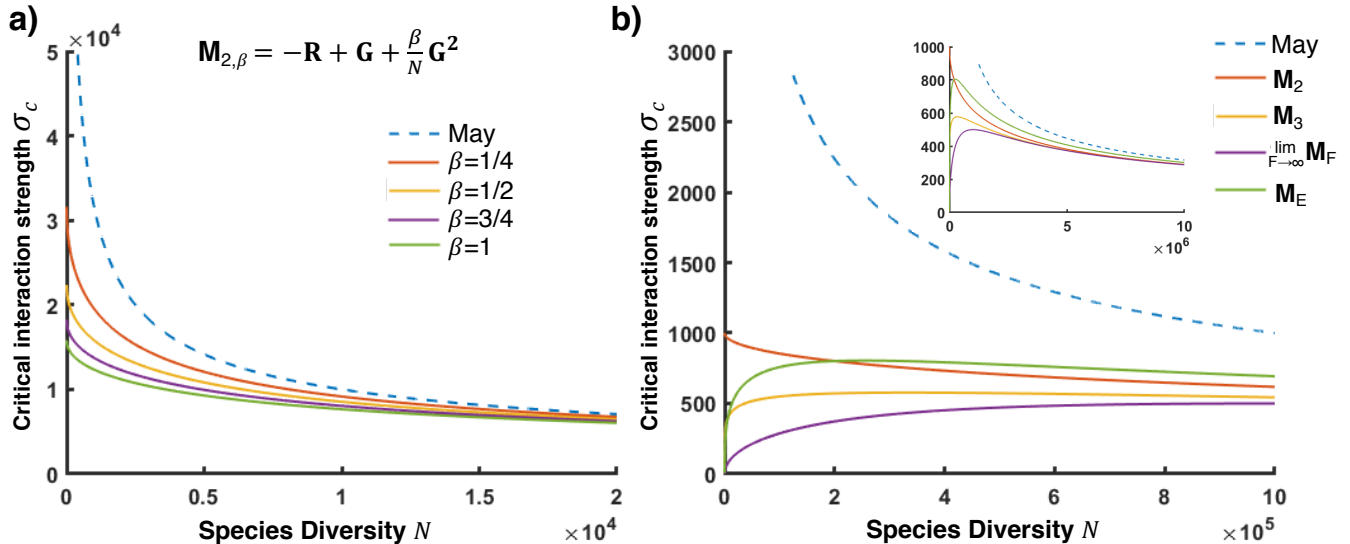
We now look at the community matrix containing up to even higher-order correlations, where we take  $\lim F \rightarrow \infty$  in  $\mathbf{M}_F$ . If we assume  $\sigma_c < \sqrt{N}$ , using the Taylor series for  $(1-x)^{-1}$  we easily find  $\sigma_c = \frac{\sqrt{NR}}{N+R}$  (see Appendix). This limit exhibits a broader peak for  $\sigma_c$  than in the  $\mathbf{M}_3$  case (Fig. 2b). One may be further interested in modifying the community matrix  $\mathbf{M}_F$ , such as with certain prefactors to the terms. A simple case is when we penalize each  $\frac{1}{N^{p-1}}\mathbf{G}^p$  term  $(p!)^{-1}$  fold, resulting in the community matrix  $\mathbf{M}_E = -\mathbf{R} + \mathbf{G} + \sum_{p=2}^F \frac{1}{N^{p-1}} \frac{\mathbf{G}^p}{p!}$ . Here, using the Taylor series for the matrix exponential of  $\mathbf{G}$ , it is straightforward to find  $\sigma_c = \sqrt{N} \log(1 + \frac{R}{N})$  for  $\mathbf{M}_E$  (Fig. 2b; see Appendix).

## Section II: negative 3-species correlations in predator-prey, mutualistic and competitive networks

We move on to analyze a different correlation structure. First, instead of May's community matrix  $\mathbf{M}_1 = -\mathbf{R} + \mathbf{G}$ , following Allesina and Tang [14], we start with a more general community matrix,  $\mathbf{M}_c = -\mathbf{R} + \mathbf{G}_c$ . The  $N \times N$  matrix  $\mathbf{G}_c$  has pairwise correlation  $\rho$  among its diagonally symmetric entries  $g_{ij}$  and  $g_{ji}$  ( $i \neq j$ ). In the construction of  $\mathbf{G}_c$ , its entries  $g_{ij}$  and  $g_{ji}$  ( $i \neq j$ ) are sampled from a general bivariate distribution with correlation  $-1 < \rho < 1$ , marginal means 0 and marginal standard deviation (interaction strength)  $\sigma$ . When  $\rho < 0$ ,  $\mathbf{M}_c$  is biased towards having more predator-prey-like ( $g_{ij}g_{ji} < 0$ ) interactions. When  $\rho > 0$ ,  $\mathbf{M}_c$  is instead biased towards having more mutualistic ( $g_{ij} > 0$ ;  $g_{ji} > 0$ ) and competitive ( $g_{ij} < 0$ ;  $g_{ji} < 0$ ) interactions which we call mutualist-competitive biased. The case of  $\rho = 0$  reduces to May's model.

To investigate the stability of the ecosystem governed by the community matrix  $\mathbf{M}_c$ , we need to determine its largest real eigenvalue components. This was done by Allesina and Tang [14] with the help of preceding theoretical results [48]. The eigenvalues of  $\mathbf{G}_c$  were shown to be distributed uniformly inside an ellipse centered at the origin in the complex plane, where the ellipse intercepts the real axis at  $\pm\sigma\sqrt{N}(1+\rho)$  and the imaginary axis at  $\pm\sigma\sqrt{N}(1-\rho)$ . The stability criterion for  $\mathbf{M}_c$  is thus  $\sigma\sqrt{N}(1+\rho) < R$  and the critical interaction strength  $\sigma_c = \frac{R}{\sqrt{N}(1+\rho)}$ . This analysis reveals a simple relationship where predator-prey biased interactions provide more stability than

FIG. 2. Non-monotonic stability relationships arise from beyond 3-species correlations



(a) Increasing  $\beta$  decreases the critical interaction strength  $\sigma_c$  of community matrix  $\mathbf{M}_{2,\beta}$ . As  $\beta$  approaches 0,  $\sigma_c$  for May's community matrix  $\mathbf{M}_1$  is recovered (dashed blue). (b) The critical interaction strengths of  $\mathbf{M}_2$  and May's community matrix  $\mathbf{M}_1$  monotonically decrease, while the community matrices including beyond 3-species correlations have a non-monotonic relationship between their critical interaction strength and species diversity. The inset shows the same plot with different scaling of the axes.  $R$  is set to  $10^6$  for all curves.

mutualist-competitive biased interactions.

We will now modify  $\mathbf{M}_c$  to introduce a different form of 3-species correlations than we studied before, which is the addition of  $-\frac{1}{N}\mathbf{G}_c^2$  to  $\mathbf{M}_c$  where we note the choice of the minus sign. This choice is partially due to the rich stability behavior it will be shown to generate. We will thus focus on the stability of the ecosystem governed by the community matrix

$$\mathbf{M}_{c,-2} = -\mathbf{R} + \mathbf{G}_c - \frac{1}{N}\mathbf{G}_c^2 \quad (4)$$

Following a similar procedure to the previous section leads to a piecewise analytical solution for the critical interaction strength  $\sigma_c$  for  $\mathbf{M}_{c,-2}$  (Fig. 3; see Appendix). An important feature of this solution is that as the species diversity decreases, including more positive as opposed to more negative correlations will transition from decreasing  $\sigma_c$  to increasing  $\sigma_c$  (Fig. 3b). This is as opposed to  $\mathbf{M}_c$  where decreasing  $\rho$  increases  $\sigma_c$  for all  $N$  (Fig. 3a).

## DISCUSSION

### Summary

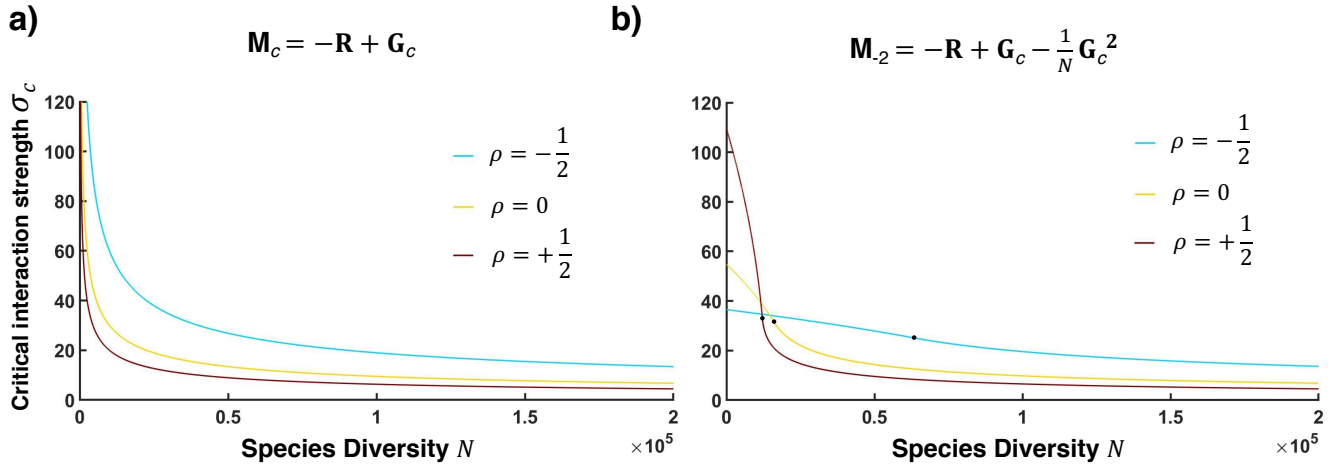
In this manuscript, we modified May's random community matrix model [7] with simple high-order correlations among species interactions by using random matrices raised to integer powers. We showed in

Section I that high-order correlations can lead, in a number of ways, to maximum possible critical interaction strengths occurring at finite values of species diversity. One implication of this is that, unlike May's model, systems with fewer species can be more unstable. A similar effect has also been observed when high-order species interactions, instead of correlations, were included in May-like models [10]. In Section II, we studied a 3-species correlation structure which resulted in predator-prey networks transitioning to being more stable than mutualist-competitive networks at higher species diversity. Our work adds to the limited number of analytical studies that involve application of random matrices with high-order correlations among their elements [14, 40, 49].

### Motivating the models

It is interesting to think about how high-order correlations in species interactions may arise in Nature. Lions, Hyenas and Cheetahs, who compete for many prey species, actively kill each others' cubs, suggestive of similarities with high-order (3-species) negative correlations as in Section I. To illustrate how parts of our models could potentially map to such biological settings, we provide a few speculative examples. Let us focus on a hypothetical scenario inspired by the interaction between Lions and Hyenas, and show one specific way increased inter-species aggression may come about in an ecosystem

FIG. 3. Mutualist-competitive and predator-prey correlations shape effect of a 3-species correlation structure



(a) Critical interaction strength  $\sigma_c$  for community matrix  $\mathbf{M}_c$  decreases with increasing  $\rho$  for all  $N$ . (b)  $\sigma_c$  for community matrix  $\mathbf{M}_{c,-2}$  is described by a piecewise function, with the two pieces joining at the respective threshold species diversity for the given curve, marked with a black dot. For  $\mathbf{M}_{c,-2}$ , increasing  $\rho$  increases  $\sigma_c$  for small  $N$  while it decreases  $\sigma_c$  for large  $N$ .  $R$  is set to 3000 for all curves.

where high-order correlations are developing in a manner consistent with our models. What we will effectively be doing is restating the mathematics of our model in words. The species names are mostly symbolic for our purposes. Interspecies aggression is known to have evolved between Lions and Hyenas. Let us assume that they exist in an ecosystem with some “initial” community matrix  $\mathbf{Q}$  (where we will ignore the self-regulation term), and that Lions eat prey A ( $q_{LA} > 0$ ;  $q_{AL} < 0$  where  $L$  is the index for Lion,  $A$  is the index for prey A,  $q$  is an element of the matrix  $\mathbf{Q}$ ) and Hyenas also eat prey A ( $q_{HA} > 0$ ;  $q_{AH} < 0$  where  $H$  is the index for Hyenas). The Lions and Hyenas also exhibit some initial, but small level of direct interaction, such that  $q_{HL}$  and  $q_{LH}$  are approximately zero.

We are interested in high-order correlations in an ecological network. High-order correlations could in theory come about in numerous ways – for example new interactions may simultaneously emerge through evolution and change predator-prey and predator-predator interactions altogether to new “values” that are unrelated to the pre-evolved values of the interactions that lacked high-order correlations. Or high-order correlations may have already been part of the ecosystem from its original formation and could continue remaining within an ecosystem, as the ecosystem evolves.

However, perhaps an alternative attractive possibility is that ecological networks vary in their level of interspecies interaction correlation over time, and correlations arise due to previous (pre-evolved) interactions feeding into how the new interactions evolve. So in the Lion and Hyena example, one way this could happen is if  $q_{LA}$  and  $q_{AH}$  feed into the evolved  $k_{LH}$  (where  $k$  is an element of the newly evolved matrix  $\mathbf{K}$ ) in this new, evolved ecological network, and likewise for  $q_{HA}$ ,  $q_{AL}$  and  $k_{HL}$ . A simple model would be to take the new  $k_{LH}$  to be proportional to  $q_{LA}q_{AH}$  — that is, the interaction of each with their common prey feeds into the interaction they are developing, and it does so in a directly proportional manner. If there are multiple common prey species (prey A, B, C, D), a simple model would be to take the average of  $q_{LA}q_{AH}$ ,  $q_{LB}q_{BH}$ ,  $q_{LC}q_{CH}$ ,  $q_{LD}q_{DH}$ . At the same time,  $k_{HL}$  would also evolve in an analogous manner, in addition to interactions of other pairs of species. Thus, this reasoning shows how one can arrive to a term similar to the addition of  $\frac{1}{N}\mathbf{Q}^2$  in our random matrix models (Figure 1b). If the mechanism described here for the arising of correlations repeats over time as the ecosystem (and consequently its community matrix) evolves further, higher powers of  $\mathbf{Q}$  than 2 will also arise. This can be seen through the following

logic: say the evolved ecosystem with correlations as above has the community matrix  $\alpha\mathbf{Q} + \beta\mathbf{Q}^2$  (where  $\alpha$ ,  $\beta$  are arbitrary constants), as opposed to the original starting ecosystem with interaction matrix  $\alpha\mathbf{Q}$  (the addition of  $\beta\mathbf{Q}^2$  is responsible for the already evolved correlations). If this evolved ecosystem with community matrix  $\alpha\mathbf{Q} + \beta\mathbf{Q}^2$  undergoes another “step” of evolution to generate correlations (just like the starting interaction matrix did), the new community matrix in this second step of evolution will include a term proportional to  $(\alpha\mathbf{Q} + \beta\mathbf{Q}^2)^2$ , and thus will include higher powers of  $\mathbf{Q}$ , namely,  $\mathbf{Q}^3$  and  $\mathbf{Q}^4$ . Mechanistically, when considering the evolution of interspecies aggression among predators, a high-order correlation, it is interesting to consider the historical selection pressure of indirect interactions (e.g. depletion of the common prey population by the other predator) and high-order interactions (e.g. driven by physical encounters of the common prey, hyena and lion). Similar arguments can be made for potential prey species that share a common predator and need to compete for sheltering territory (a high-order interaction), which may be expected to later correlate with evolution of interspecies aggression among the prey species. However, whether and how our models may apply to the full variety of species interactions in large ecosystems is an interesting avenue for future investigation.

We would like to say that we believe the “mappings” in this section are extremely speculative, despite being mathematically interesting. The hypothetical Lion-Hyena example helps motivate the correlation structure in Section I of our manuscript, and in particular the positive sign in front of  $\mathbf{G}^2$  as well as the integer powers of our matrices. Overall, it would be interesting to study the inter-species aggression behavior of other higher predators that occur independent of the abundance of their common prey species, and look for patterns within ecosystems across many predators for high-order negative correlations in species interactions, and their effects on ecological stability (see patterns discussed in [50]).

On the other hand, ancient Humans and ancient Dogs who shared many prey species likely hunted together: a high-order interaction, as opposed to a high-order correlation in their direct interaction, since it depends on presence of another species. However, they later seem to have developed a direct interaction of companionship independent of their shared prey; such companionship among predators could be an interesting angle to investigate high-order positive correlations as in Section II. Thus, in order to motivate the matrix structure in Section II, we would like to focus on another hypothetical scenario inspired by the interaction of ancient Humans and Dogs, which again is meant to be mostly symbolic. We will take Rabbits as their common prey. Let us assume ancient Humans fed on Rabbits ( $q_{HR} > 0$ ,  $q_{RH} < 0$ ) and Dogs fed on Rabbits ( $q_{DR} > 0$ ,  $q_{RD} < 0$ ). Then, cooperative hunting evolved (a high-order interaction, not correlation – Humans hunted Rabbits more efficiently

in presence of Dogs). However, Humans then over time also evolved a direct companionship with Dogs, where even in the absence of Rabbits, Humans provided shelter for Dogs. Here, if we take the same approach as the Lion and Hyena model to calculate the newly evolved  $k_{HD}$ , we see that  $q_{HR}q_{RD}$  would be negative. Thus, for a direct favorable interaction between Humans and Dogs, we take  $k_{HD}$  to be proportional to  $-q_{HR}q_{RD}$ . This motivates the negative sign in front of  $\mathbf{G}^2$  in Section II.

In this hypothetical ancient Human-Dog example, we reasoned that the evolution of high-order interactions (e.g. cooperative hunting) could lead to evolution of high-order correlations in direct interactions. We are not aware of any conclusive evidence for this being a realistic mechanism, but it seems like an attractive possibility.

### Future avenues

We dropped the connectivity fraction from May’s model, effectively setting it to 1, since it becomes more subtle in our framework—the connectivity in general increases as we raise our matrices to different powers, which introduces interactions among otherwise non-interacting (disconnected) species. A future avenue for investigation would thus be the implications of different levels as well as patterns of connectivity in our community matrices and the matrices that are used to construct them. Furthermore, it would be interesting to study the introduction of matrices with high-order correlations as community matrices into generalized Lotka-Volterra models. Likewise, comparing our matrices with steady state Jacobian matrices from different models of high-order interactions should also be informative.

Importantly, as mentioned before, future investigations are required to understand whether and how our models make sense when the full variety of species interactions in a large ecosystem are considered. We also acknowledge that even if high-order correlations do arise in a qualitatively similar fashion to our models, the appropriate quantitative description for them may still be different. For example, the proportionality we describe above as well as the way we do our averaging with  $N$  (for example the  $\frac{1}{N}$  factor in front of  $\mathbf{G}^2$ ) are, after all, particular choices in our model that made sense to us (see Figure 1b and description in Section I). Additionally, while it seems likely that high-order correlations among species interactions do exist in some form in Nature, it is of course unknown whether high-order correlations in practice even produce any of the various ecological stability behaviors that we predict. So our model is a very speculative one. Additionally, our models are phenomenological in nature, just like many other ecological models, and thus mechanistic studies (e.g. with evolutionary game theory) are required to check if they at all make sense in the context of evolution.

Although our models can perhaps be modified



in the future, in their current form they involve introducing high-order correlations throughout the ecological network for all species, instead of for a few select group of species. In this light, it is interesting to consider the case of Amazonian birds, where many species of birds sharing similar resources, as opposed to just a few pairs, seem to have developed interspecies aggression [51] where they fight, for example, over territory. We think these kinds of settings with many species being involved in interspecies aggression could overall be fitting starting points to study the effect of high-order correlations on stability.

An important orthogonal avenue for consideration for biological relevance of our results is when whole ecosystem properties are under effective selective pressure. Typically, each species within an ecosystem is considered to be individually under selection. However, additional effective selection pressures on whole ecosystems can potentially be the case in microbiota attached to hosts (where the microbiota are taken to be ecosystems on their own): here, selective pressures may perhaps act on not only individual microbial species, but also on the microbial ecosystem as a whole, through outputs of the ecosystem [5]. Thus, for example, if stability of the microbial ecosystem is a favorable aspect to the host, our frameworks in Section I and Section II show recipes for achieving such favorable regimes, that could thus emerge due to selection, because of the advantage provided to the host, whatever the mechanistic basis. However, this argument would require evolutionary considerations, such as evolutionary game theory, to test its merits.

Perhaps a particularly important use for our framework could be in reverse, that is to utilize the forms of correlation structures we propose to achieve stability in engineered ecosystems. While this seems out of reach with current technology, microbiota engineering is a rapidly developing field [52]. It seems reasonable to think that having stable microbiota in the human body could be a desirable feature. As bioengineering progresses, perhaps an improved version of the framework we propose could be utilized to help determine what modifications to make in human microbiota to make it more stable. Our framework is potentially attractive

in this sense because it gives a recipe for modifying the stability of the ecosystem by introducing the same operation to each species interaction within the ecosystem, instead of having to specifically target few strong/weak species or perhaps having to know details of the species. It is however unclear to us what physical form (e.g. details of what should be included in a potential plasmid library) such a perturbation would take.

Overall, like many ecological models, ours are phenomenological and await consolidation and complementation via input from metabolics, genetics as well as branches of evolutionary theory [5, 53–55]. Understanding finite size effects in our models is another avenue for future investigation [56, 57]. We hope that as experimental challenges in studying species interactions are resolved [6, 43, 58, 59], for example by using microbial communities as models of large ecosystems [5, 60–65], our theoretical predictions will be rigorously tested.

## ACKNOWLEDGEMENTS

E.K. is supported by the Stanford Bio-X Graduate Fellowship and the Biophysics Ph.D. Program at Stanford University. A.R.D. acknowledges the HHMI (Faculty Scholar Award), and the NIH (R35GM130332). We thank Dr. William Gilpin, Dr. Po-Yi Ho, Samuel Bray, Dr. Bo Wang and Dr. Kerwyn Casey Huang for discussion and feedback on the manuscript. We thank the three anonymous reviewers and the editor whose suggestions helped improve the manuscript.

## AUTHOR CONTRIBUTIONS

E.K. conceived the idea, conducted the research and wrote the manuscript. E.K. and A.R.D. discussed and prepared the manuscript.

## ADDITIONAL INFORMATION

**The authors declare no competing interests.**

- 
- [1] F. M. Guedes, G. S. Garcia, L. M. Versieux, L. Q. Matias, and M. Alves, Insights on underestimated Lentibulariaceae diversity in northeastern Brazil: new records and notes on distribution, diversity and endemism in the family, *Brazilian Journal of Botany* **41**, 867 (2018).
- [2] S. A. Karl, A. L. F. Castro, J. A. Lopez, P. Charvet, and G. H. Burgess, Phylogeography and conservation of the bull shark (*Carcharhinus leucas*) inferred from mitochondrial and microsatellite DNA, *Conservation*

- Genetics* **12**, 371 (2011).
- [3] K. Z. Coyte and S. Rakoff-Nahoum, Understanding Competition and Cooperation within the Mammalian Gut Microbiome, *Current Biology* **29**, R538 (2019).
- [4] J. Gilbert, M. J. Blaser, J. G. Caporaso, J. Jansson, S. V. Lynch, and R. Knight, Current understanding of the human microbiome, *Nature medicine* **24**, 392 (2018).
- [5] K. R. Foster, J. Schluter, K. Z. Coyte, and S. Rakoff-Nahoum, The evolution of the host microbiome as an ecosystem on a leash, *Nature* **548**, 43 (2017), number:

- 7665 Publisher: Nature Publishing Group.
- [6] S. Widder, R. J. Allen, T. Pfeiffer, T. P. Curtis, C. Wiuf, W. T. Sloan, O. X. Cordero, S. P. Brown, B. Momeni, W. Shou, H. Kettle, H. J. Flint, A. F. Haas, B. Laroche, J.-U. Kreft, P. B. Rainey, S. Freilich, S. Schuster, K. Milferstedt, J. R. van der Meer, T. Gropkopf, J. Huisman, A. Free, C. Picioreanu, C. Quince, I. Klapper, S. Labarthe, B. F. Smets, H. Wang, and O. S. Soyer, Challenges in microbial ecology: building predictive understanding of community function and dynamics, *The ISME Journal* **10**, 2557 (2016), number: 11 Publisher: Nature Publishing Group.
- [7] R. M. May, Will a Large Complex System be Stable?, *Nature* **238**, 413 (1972), number: 5364 Publisher: Nature Publishing Group.
- [8] S. Allesina and S. Tang, The stability–complexity relationship at age 40: a random matrix perspective, *Population Ecology* **57**, 63 (2015).
- [9] T. Namba, Multi-faceted approaches toward unravelling complex ecological networks, *Population Ecology* **57**, 3 (2015).
- [10] E. Bairey, E. D. Kelsic, and R. Kishony, High-order species interactions shape ecosystem diversity, *Nature Communications* **7**, 10.1038/ncomms12285 (2016).
- [11] J. Ginibre, Statistical Ensembles of Complex, Quaternion, and Real Matrices, *Journal of Mathematical Physics* **6**, 440 (1965), publisher: American Institute of Physics.
- [12] V. L. Girko, Circular Law, *Theory of Probability & Its Applications* **29**, 694 (1985), publisher: Society for Industrial and Applied Mathematics.
- [13] T. Tao, V. Vu, and M. Krishnapur, Random matrices: Universality of ESDs and the circular law, *Annals of Probability* **38**, 2023 (2010), publisher: Institute of Mathematical Statistics.
- [14] S. Allesina and S. Tang, Stability criteria for complex ecosystems, *Nature* **483**, 205 (2012), number: 7388 Publisher: Nature Publishing Group.
- [15] M. Novak, J. D. Yeakel, A. E. Noble, D. F. Doak, M. Emmerson, J. A. Estes, U. Jacob, M. T. Tinker, and J. T. Wootton, Characterizing Species Interactions to Understand Press Perturbations: What Is the Community Matrix?, *Annual Review of Ecology, Evolution, and Systematics* **47**, 409 (2016), publisher: Annual Reviews.
- [16] H.-C. Ho, J. M. Tylianakis, and S. Pawar, Behaviour moderates the impacts of food-web structure on species coexistence, *Ecology Letters* **n/a**, <https://doi.org/10.1111/ele.13643>, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ele.13643>.
- [17] L. Stone, The feasibility and stability of large complex biological networks: a random matrix approach, *Scientific Reports* **8**, 8246 (2018), number: 1 Publisher: Nature Publishing Group.
- [18] M. Dougoud, L. Vinckenbosch, R. P. Rohr, L.-F. Bersier, and C. Mazza, The feasibility of equilibria in large ecosystems: A primary but neglected concept in the complexity-stability debate, *PLOS Computational Biology* **14**, e1005988 (2018), publisher: Public Library of Science.
- [19] J. Grilli, M. Adorisio, S. Suweis, G. Barabás, J. R. Banavar, S. Allesina, and A. Maritan, Feasibility and coexistence of large ecological communities, *Nature Communications* **8**, 14389 (2017), number: 1 Publisher: Nature Publishing Group.
- [20] L. Stone, The Google matrix controls the stability of structured ecological and biological networks, *Nature Communications* **7**, 12857 (2016), number: 1 Publisher: Nature Publishing Group.
- [21] P. Landi, H. O. Minoarivelo, Brännström, C. Hui, and U. Dieckmann, Complexity and stability of ecological networks: a review of the theory, *Population Ecology* **60**, 319 (2018).
- [22] Y. V. Fyodorov and B. A. Khoruzhenko, Nonlinear analogue of the May-Wigner instability transition, *Proceedings of the National Academy of Sciences* **113**, 6827 (2016), publisher: National Academy of Sciences Section: Physical Sciences.
- [23] A. Mougi and M. Kondoh, Diversity of Interaction Types and Ecological Community Stability, *Science* **337**, 349 (2012), publisher: American Association for the Advancement of Science Section: Report.
- [24] C. Jacquet, C. Moritz, L. Morissette, P. Legagneux, F. Massol, P. Archambault, and D. Gravel, No complexity–stability relationship in empirical ecosystems, *Nature Communications* **7**, 12573 (2016), number: 1 Publisher: Nature Publishing Group.
- [25] S. Allesina, J. Grilli, G. Barabás, S. Tang, J. Aljadeff, and A. Maritan, Predicting the stability of large structured food webs, *Nature Communications* **6**, 10.1038/ncomms8842 (2015).
- [26] K. Z. Coyte, J. Schluter, and K. R. Foster, The ecology of the microbiome: Networks, competition, and stability, *Science* **350**, 663 (2015), publisher: American Association for the Advancement of Science Section: Report.
- [27] G. D. Kokkoris, V. A. A. Jansen, M. Loreau, and A. Y. Troumbis, Variability in interaction strength and implications for biodiversity, *Journal of Animal Ecology* **71**, 362 (2002).
- [28] S. Tang, S. Pawar, and S. Allesina, Correlation between interaction strengths drives stability in large ecological networks, *Ecology Letters* **17**, 1094 (2014), eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ele.12312>.
- [29] D. Gravel, F. Massol, and M. A. Leibold, Stability and complexity in model meta-ecosystems, *Nature Communications* **7**, 12457 (2016), number: 1 Publisher: Nature Publishing Group.
- [30] T. Okuyama and J. N. Holland, Network structural properties mediate the stability of mutualistic communities, *Ecology Letters* **11**, 208 (2008), eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1461-0248.2007.01137.x>.
- [31] M. T. Pearce, A. Agarwala, and D. S. Fisher, Stabilization of extensive fine-scale diversity by ecologically driven spatiotemporal chaos, *Proceedings of the National Academy of Sciences* **117**, 14572 (2020), publisher: National Academy of Sciences Section: Biological Sciences.
- [32] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, Random matrix approach to cross correlations in financial data, *Physical Review E* **65**, 066126 (2002), publisher: American Physical Society.
- [33] Y. Bahri, J. Kadmon, J. Pennington, S. S. Schoenholz, J. Sohl-Dickstein, and S. Ganguli, Statistical Mechanics of Deep Learning, *Annual Review of Condensed Matter Physics* **11**, 501 (2020), eprint:

- <https://doi.org/10.1146/annurev-conmatphys-031119-050745>.
- [34] M. L. Manning and A. J. Liu, A random matrix definition of the boson peak, *EPL (Europhysics Letters)* **109**, 36002 (2015), publisher: IOP Publishing.
- [35] A. Edelman and N. R. Rao, Random matrix theory, *Acta Numerica* **14**, 233 (2005), publisher: Cambridge University Press.
- [36] A. Edelman and B. D. Sutton, From Random Matrices to Stochastic Operators, *Journal of Statistical Physics* **127**, 1121 (2007).
- [37] A. Edelman and Y. Wang, Random Matrix Theory and Its Innovative Applications, in *Advances in Applied Mathematics, Modeling, and Computational Science*, Fields Institute Communications, edited by R. Melnik and I. S. Kotsireas (Springer US, Boston, MA, 2013) pp. 91–116.
- [38] S. Franz, G. Parisi, P. Urbani, and F. Zamponi, Universal spectrum of normal modes in low-temperature glasses, *Proceedings of the National Academy of Sciences* **112**, 14539 (2015), publisher: National Academy of Sciences Section: Physical Sciences.
- [39] T. Guhr, A. Müller–Groeling, and H. A. Weidenmüller, Random-matrix theories in quantum physics: common concepts, *Physics Reports* **299**, 189 (1998).
- [40] D. S. Dean and F. Ritort, Squared interaction matrix Sherrington–Kirkpatrick model for a spin glass, *Physical Review B* **65**, 224209 (2002), publisher: American Physical Society.
- [41] C. Tu, S. Suweis, J. Grilli, M. Formentin, and A. Maritan, Reconciling cooperation, biodiversity and stability in complex ecological communities, *Scientific Reports* **9**, 5580 (2019), number: 1 Publisher: Nature Publishing Group.
- [42] J. T. Wootton, The Nature and Consequences of Indirect Effects in Ecological Communities, *Annual Review of Ecology and Systematics* **25**, 443 (1994), eprint: <https://doi.org/10.1146/annurev.es.25.110194.002303>.
- [43] J. T. Wootton and M. Emmerson, Measurement of Interaction Strength in Nature, *Annual Review of Ecology, Evolution, and Systematics* **36**, 419 (2005), eprint: <https://doi.org/10.1146/annurev.ecolsys.36.091704.175535>.
- [44] J. Grilli, G. Barabás, M. J. Michalska-Smith, and S. Allesina, Higher-order interactions stabilize dynamics in competitive network models, *Nature* **548**, 210 (2017), number: 7666 Publisher: Nature Publishing Group.
- [45] J. C. D. Terry, R. J. Morris, and M. B. Bonsall, Interaction modifications lead to greater robustness than pairwise non-trophic effects in food webs, *Journal of Animal Ecology* **88**, 1732 (2019), eprint: <https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/1365-2656.13057>.
- [46] H. Mickalide and S. Kuehn, Higher-Order Interaction between Species Inhibits Bacterial Invasion of a Phototroph-Predator Microbial Community, *Cell Systems* **9**, 521 (2019).
- [47] A. Sanchez, Defining higher-order interactions in synthetic ecology: lessons from physics and quantitative genetics, *Cell systems* **9**, 519 (2019).
- [48] H. J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein, Spectrum of Large Random Asymmetric Matrices, *Physical Review Letters* **60**, 1895 (1988), publisher: American Physical Society.
- [49] P. V. Aceituno, T. Rogers, and H. Schomerus, Universal hypotrochoidic law for random matrices with cyclic correlations, *Physical Review E* **100**, 010302 (2019), publisher: American Physical Society.
- [50] F. Palomares and T. M. Caro, Interspecific Killing among Mammalian Carnivores., *The American Naturalist* **153**, 492 (1999), publisher: The University of Chicago Press.
- [51] S. K. Robinson and J. Terborgh, Interspecific Aggression and Habitat Selection by Amazonian Birds, *Journal of Animal Ecology* **64**, 1 (1995), publisher: [Wiley, British Ecological Society].
- [52] C.-Y. Chang, J. C. C. Vila, M. Bender, R. Li, M. C. Mankowski, M. Bassette, J. Borden, S. Golfier, P. G. L. Sanchez, R. Waymack, X. Zhu, J. Diaz-Colunga, S. Estrela, M. Rebolleda-Gomez, and A. Sanchez, Engineering complex communities by directed evolution, *Nature Ecology & Evolution*, 1 (2021), publisher: Nature Publishing Group.
- [53] J. N. Thompson, The Evolution of Species Interactions, *Science* **284**, 2116 (1999), publisher: American Association for the Advancement of Science Section: Viewpoint.
- [54] S. K. Hansen, P. B. Rainey, J. A. J. Haagensen, and S. Molin, Evolution of species interactions in a biofilm community, *Nature* **445**, 533 (2007), number: 7127 Publisher: Nature Publishing Group.
- [55] R. Allen and B. Waclaw, Antibiotic resistance: a physicist’s view, *Physical Biology* **13**, 10.1088/1478-3975/13/4/045001 (2016), publisher: IOP Publishing.
- [56] C. A. Tracy and H. Widom, The Distributions of Random Matrix Theory and their Applications, in *New Trends in Mathematical Physics*, edited by V. Sidoravičius (Springer Netherlands, Dordrecht, 2009) pp. 753–765.
- [57] Z. Bai and J. W. Silverstein, Convergence Rates of ESD, in *Spectral Analysis of Large Dimensional Random Matrices*, Springer Series in Statistics, edited by Z. Bai and J. W. Silverstein (Springer, New York, NY, 2010) pp. 181–221.
- [58] F. Carrara, A. Giometto, M. Seymour, A. Rinaldo, and F. Altermatt, Inferring species interactions in ecological communities: a comparison of methods at different levels of complexity, *Methods in Ecology and Evolution* **6**, 895 (2015).
- [59] D. J. Harris, Inferring species interactions from co-occurrence data with Markov networks, *Ecology* **97**, 3308 (2016).
- [60] C. D. Nadell, K. Drescher, and K. R. Foster, Spatial structure, cooperation and competition in biofilms, *Nature Reviews Microbiology* **14**, 589 (2016), number: 9 Publisher: Nature Publishing Group.
- [61] R. R. Stein, V. Bucci, N. C. Toussaint, C. G. Buffie, G. Rättsch, E. G. Pamer, C. Sander, and J. B. Xavier, Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota, *PLOS Computational Biology* **9**, e1003388 (2013), publisher: Public Library of Science.
- [62] C. Ratzke, J. Barrere, and J. Gore, Strength of species interactions determines biodiversity and stability in microbial communities, *Nature Ecology & Evolution* **4**, 376 (2020), number: 3 Publisher: Nature Publishing Group.
- [63] A. L. Goodman, N. P. McNulty, Y. Zhao, D. Leip, R. D. Mitra, C. A. Lozupone, R. Knight, and J. I. Gordon, Identifying Genetic Determinants Needed to Establish

a Human Gut Symbiont in Its Habitat, Cell Host & Microbe **6**, 279 (2009).

- [64] K. Foster and T. Bell, Competition, Not Cooperation, Dominates Interactions among Culturable Microbial Species, Current Biology **22**, 1845 (2012).
- [65] K. Coyte, C. Rao, S. Rakoff-Nahoum, and K. Foster, Community assembly in the microbiome: ecological insights into infant microbiome development, Access Microbiology, **1**, 241 (2019), publisher: Microbiology Society,.
- [66] D. Sherrington and S. Kirkpatrick, Solvable Model of a Spin-Glass, Physical Review Letters **35**, 1792 (1975), publisher: American Physical Society.

## APPENDIX

### Squared interaction matrix Sherrington-Kirkpatrick model

The Sherrington-Kirkpatrick spin-glass model is an Ising model where infinite-ranged random interactions couple the spins, in the absence of an external field [66]. The random spin-spin interaction matrix  $\mathbf{J}$  (the Hessian of the Ising Hamiltonian) has entries that are independent and identically Gaussian distributed. In order to study next-to-nearest neighbor correlations, Dean and Ritort [40] replaced the interaction matrix in the Sherrington-Kirkpatrick model with a “squared interaction matrix”, namely  $\mathbf{J}\mathbf{J}^T$ .

#### Section I details

When we refer to the eigenvalue distribution of a finite-sized random matrix, we consider the approximation of the ensemble average eigenvalue distribution from many instantiations of the matrix. For large  $N$  [13], the largest eigenvalue real part in the eigenvalue distribution of  $\mathbf{M}_{2,\beta} = -\mathbf{R} + \mathbf{G} + \frac{\beta}{N}\mathbf{G}^2$ , where  $\mathbf{G}$  is defined as in the manuscript, can be found as follows. Note that the last 2 terms share the same set of eigenvectors, which are also eigenvectors of  $\mathbf{R} = R\mathbf{I}$  where  $\mathbf{I}$  is the identity matrix. In general, if  $v$  is an eigenvalue of  $\mathbf{G}$ , and  $\chi$  is a polynomial,  $\chi(v)$  is an eigenvalue of  $\chi(\mathbf{G})$ . Thus the largest eigenvalue real part in the eigenvalue distribution of  $\mathbf{M}_{2,\beta}$  approaches  $-R + \sigma\sqrt{N} + \beta\sigma^2$  for large  $N$ . To find the critical interaction strength  $\sigma_c$  above which the ecosystem becomes unstable, we equate this limiting largest eigenvalue real part to 0 and solve for  $\sigma_c$ , which gives  $\sigma_c = \frac{1}{2\beta}(\sqrt{4\beta R + N} - \sqrt{N})$  for  $\mathbf{M}_{2,\beta}$ .

We will approach the community matrix

$$\mathbf{M}_F = -\mathbf{R} + \mathbf{G} + \sum_{p=2}^F \frac{1}{N^{p-1}}\mathbf{G}^p \quad (5)$$

similarly. For  $F = 3$ , that is  $\mathbf{M}_3$ , the largest eigenvalue real part approaches  $-R + \sigma\sqrt{N} + \sigma^2 + \frac{\sigma^3}{\sqrt{N}}$  for large  $N$ .

Equating this to 0 and solving for the critical interaction strength gives

$$\sigma_c = \frac{1}{6}\sqrt{N} \left( 2^{2/3} \sqrt[3]{\frac{27R}{N} + \sqrt{\left(\frac{27R}{N} + 7\right)^2 + 32} + 7} - \frac{4\sqrt[3]{2}}{\sqrt[3]{\frac{27R}{N} + \sqrt{\left(\frac{27R}{N} + 7\right)^2 + 32} + 7}} - 2 \right) \quad (6)$$

In the case of  $\lim F \rightarrow \infty$ , we rewrite  $\mathbf{M}_F$  as  $-\mathbf{R} - N\mathbf{I} + N \sum_{p=0}^F \frac{1}{N^p}\mathbf{G}^p$ . Thus, using the Taylor series  $\frac{1}{1-x} = \lim_{F \rightarrow \infty} \sum_{p=0}^F x^p$  (convergent for  $-1 < x < 1$ ), we get the approached largest eigenvalue real part for  $\lim_{F \rightarrow \infty} \mathbf{M}_F$  as  $-R - N + N(1 - \frac{\sigma}{\sqrt{N}}\sqrt{N})^{-1}$  for large  $N$  as long as  $\sigma < \sqrt{N}$ . Under the condition  $\sigma < \sqrt{N}$ , we can solve for when the largest eigenvalue real part, for large  $N$ , is equal to 0 which gives  $\sigma_c = \frac{\sqrt{NR}}{N+R}$ .

In the case of the community matrix  $\mathbf{M}_E = -\mathbf{R} + \mathbf{G} + \sum_{p=2}^F \frac{1}{N^{p-1}}\frac{\mathbf{G}^p}{p!}$ , we can rearrange to get  $\mathbf{M}_E = -\mathbf{R} - N\mathbf{I} + N \sum_{p=0}^F \frac{1}{N^p}\frac{\mathbf{G}^p}{p!}$ . The last term is recognized as simply  $N$  times the matrix exponential of  $\mathbf{G}$ , which leads to the largest eigenvalue real part approaching  $-R - N + N \exp(\frac{\sigma_c}{\sqrt{N}}\sqrt{N})$  for large  $N$ . Thus, we can solve  $-R - N + N \exp(\frac{\sigma_c}{\sqrt{N}}\sqrt{N}) = 0$  to get  $\sigma_c = \sqrt{N} \log(1 + \frac{R}{N})$ .

#### Section II details

As described in Section I details, we again consider how points map to each other in the eigenvalue distributions. The eigenvalues of  $\mathbf{G}_c$  lie inside the ellipse  $(\frac{x}{1+\rho})^2 + (\frac{y}{1-\rho})^2 = N\sigma^2$  in the complex plane, for large  $N$ . We can calculate real part of eigenvalues of  $\mathbf{M}_{c,-2} = -\mathbf{R} + \mathbf{G}_c - \frac{1}{N}\mathbf{G}_c^2$  as follows. The eigenvalue  $x + iy$  in the eigenvalue distribution of  $\mathbf{G}_c$  will map to  $\Psi = -R + x + iy - \frac{1}{N}(x + iy)^2 = -R + x + iy - \frac{1}{N}x^2 + \frac{1}{N}y^2 - \frac{2}{N}ixy$ , the corresponding eigenvalue of  $\mathbf{M}_{c,-2}$ . At the ellipse boundary of the approached eigenvalue distribution of  $\mathbf{G}_c$  for large  $N$ ,  $y = \pm(1 - \rho)\sqrt{N\sigma^2 - (\frac{x}{1+\rho})^2}$  and the real part of  $\Psi$  becomes  $-R + x - \frac{1}{N}x^2 + \frac{1}{N}((1 - \rho)\sqrt{N\sigma^2 - (\frac{x}{1+\rho})^2})^2 = -R + (1 - \rho)^2\sigma^2 + x - (1 + \frac{(1-\rho)^2}{(1+\rho)^2})\frac{x^2}{N}$ . This is maximized when  $1 - (1 + \frac{(1-\rho)^2}{(1+\rho)^2})\frac{x}{N} = 0$  giving  $x_{max} = \frac{N}{2(1 + \frac{(1-\rho)^2}{(1+\rho)^2})}$ . However, the range of possible  $x$  is restricted to  $-(1 + \rho)\sigma\sqrt{N} \leq x \leq (1 + \rho)\sigma\sqrt{N}$ , so when  $x_{max} > (1 + \rho)\sigma\sqrt{N}$ , that is  $\frac{1}{(1+\rho)}\frac{\sqrt{N}}{2(1 + \frac{(1-\rho)^2}{(1+\rho)^2})} > \sigma$ , the maximization within the range occurs at  $(1 + \rho)\sigma\sqrt{N}$ .

The largest eigenvalue real part, for large  $N$ , thus approaches

$$m_1 = -R + (1 - \rho)^2 \sigma^2 + \frac{\gamma N}{2}$$

for  $\sigma \geq \frac{1}{(1+\rho)} \gamma \sqrt{N}$ , and

$$m_2 = -R + (1 - \rho)^2 \sigma^2 + (1 + \rho) \sigma \sqrt{N} - \frac{1}{2\gamma} (1 + \rho)^2 \sigma^2$$

for  $\sigma < \frac{1}{(1+\rho)} \gamma \sqrt{N}$  where  $\gamma = \frac{1}{2(1+\frac{(1-\rho)^2}{(1+\rho)^2})}$ .

Thus, when  $\sigma \geq \frac{1}{(1+\rho)} \gamma \sqrt{N}$ , the ecosystem is stable under the condition  $m_1 \leq 0$ . This gives a critical interaction strength  $\sigma_c = \sqrt{(1 - \rho)^{-2} (R - \frac{\gamma N}{2})} = \frac{1}{2} \sqrt{\frac{-N+4R}{(\rho-1)^2} + \frac{N}{2(1+\rho^2)}}$  down to the threshold species diversity  $N = N_T$  where  $\sqrt{(1 - \rho)^{-2} (R - \frac{\gamma N_T}{2})} = \frac{1}{(1+\rho)} \gamma \sqrt{N_T}$ . Rearranging, we find  $N_T = (\frac{\gamma}{2} + \frac{(1-\rho)^2 \gamma^2}{(1+\rho)^2})^{-1} R = \frac{16R(\rho^2+1)^2}{(\rho+1)^2(\rho(3\rho-2)+3)}$ .

When  $\sigma < \frac{1}{(1+\rho)} \gamma \sqrt{N}$ , the ecosystem is stable as long as  $m_2 \leq 0$ . To find the critical interaction strength, we start with solving  $m_2 = 0$  which gives the curves  $\sigma = \omega_{\pm} = \frac{1}{2((1-\rho)^2 - \frac{1}{2\gamma}(1+\rho)^2)} (-(1 +$

$\rho) \sqrt{N}) \pm \sqrt{(1 + \rho)^2 N + 4R((1 - \rho)^2 - \frac{1}{2\gamma}(1 + \rho)^2)} = \frac{\sqrt{N} \mp \sqrt{(N-4R)}}{2(\rho+1)}$ . Note that when  $\sigma = \omega_-$ ,  $\sigma < \frac{1}{(1+\rho)} \gamma \sqrt{N}$  is not satisfied. Thus in this regime,  $\sigma_c = \omega_+$ .

To summarize, after rearranging and simplifications, the final piecewise function for the critical interaction strength is  $\sigma_c =$

$$\frac{1}{2} \sqrt{\frac{4R - N}{(\rho - 1)^2} + \frac{N}{2(1 + \rho^2)}}, \quad N \leq N_T$$

$$\frac{\sqrt{N} - \sqrt{(N - 4R)}}{2(\rho + 1)}, \quad N > N_T$$

for  $-1 < \rho < 1$ .

### Software

Analytical results were found by hand, using Wolfram Mathematica version 12 and WolframAlpha search engine. Plots were made using Matlab version R2019a and further labeled in Microsoft PowerPoint. Figure 1 was partially made in BioRender (biorender.com).