How collectively integrated are ecological communities?

Yuval R. Zelnik*[1], Nuria Galiana*[2], Matthieu Barbier[3], Michel Loreau[4], Eric Galbraith[5,6], and Jean-François Arnoldi†[4]

[1]Department of Crop Production Ecology, Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden
[2]Department of Biogeography and Global Change, National Museum of Natural Sciences (CSIC), 28006 Madrid, Spain.
[3]Plant Health Institute CIRAD, Montpellier, France
[4]Theoretical and Experimental Ecology Station, CNRS Moulis, France
[5]Department of Earth and Planetary Science, McGill University, Montreal, Quebec, Canada
[6]Institut de Ciència i Tecnologia Ambientals (ICTA-UAB), Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona, Spain

*Authors contributed equally
†jean-francois.arnoldi@cnrs.fr

Abstract

Are the population dynamics of species mainly determined by direct interactions with predators, preys and conspecifics? Or, instead, are those dynamics more dependent on indirect feedbacks that ripple across the whole interaction network?

Here we show that, from a basic spectral feature of the interaction network, we can predict the length of indirect interaction pathways that contribute to community-level dynamical patterns, such as the depth of a perturbation's reach, or the contribution of biotic processes to realized species niches. In doing so, we propose a measure of collectivity that integrates existing approaches to community complexity, collective integration and indirect interactions.

By revisiting classic concepts of theoretical community ecology, our work proposes an original perspective on the question of to what degree communities are more than loose collections of species or simple interaction motifs. This perspective can help clarify when reductionist approaches, focusing on particular species and small interaction motifs, ought to suffice or fail when applied to ecosystems.

Keywords: Reductionism, holism, interaction networks, indirect interactions, Lotka-Volterra models, complexity, stability, perturbations, eigenvalues.

Introduction

> ”The Scientific Revolution of the seventeenth century legitimated the idea of a mechanical nature in which the behavior of every element can be explained by laws, within a totality seen as the sum of its parts and the interactions of those elements.”
>
> Philippe Descola, *Beyond Nature and Culture.*

Ecological communities comprise vast networks of interacting species, that greatly vary in their richness and connectivity (Pimm, 1984; Montoya *et al.*, 2006; Agrawal, 2001; Brown *et al.*, 2001). To understand and predict the behaviour of ecological systems it is tempting to break them down into small parts that are individually much easier to study (predator-prey pairs, competitors within a same niche, etc.). How useful this approach can be depends on how easy it is to scale-up knowledge about individual pieces to understand features of the whole. For instance, we might hope that, to understand the population dynamics of a given species, it would be sufficient to consider only the other species with which it interacts directly, and most strongly. But this is far from self-evident. In fact, it could be just as reasonable to think that a species response to environmental change strongly depends on responses of the whole ecosystem in which it is embedded (Patten, 1982).

Explicitly or implicitly, for decades ecologist have argued about whether reductionist or holistic perspectives are most appropriate (Loreau, 2020). This contrast in approaches to ecosystems is often traced back to the opposition, regarding plant communities, between the holistic view of Clements, to the parsimonious individualistic perspective of Gleason. Clements argued that plant associations should be understood as high-level biological entities, comparable to actual organisms, so that species are best understood through their functions within a whole (Clements, 1916). Gleason claimed that plant communities are mere collections of individual species and gave little importance to the interactions between them (Gleason, 1926). This dichotomy has carried on, with notable ideas such as Lovelock's Gaia theory proposing that the biosphere should be viewed as a super-organism (Lovelock & Margulis, 1974), a perspective on ecological systems that profoundly contrasts with Hubbell's neutral theory, which explains biodiversity patterns via random effects of migration and extinctions alone (Hubbell, 2001), or Species Distribution Models that predict species ranges from a few key environmental variables (Soberón, 2007).

The epistemological difficulty lies in the fact that, by focusing on clearly identifiable individual mechanisms, individual genes, individual species, and using controlled experiments, the reductionist approach has arguably led to the greatest advances of biology. Yet, given the complexity, and interwoven levels of organization present in ecological systems, it seems overly optimistic to think that decomposing communities into small, clearly defined and well studied parts, will suffice to understand the whole (Levins & Lewontin, 1982; Lewontin, 1996; Bergelson *et al.*, 2021).

In fact, a clear challenge to the individualitic view point is the existence of many indirect interactions between species. These interactions are mediated via one or several intermediate populations (Wootton, 1994). They form long, and numerous, pathways across the interaction network (Puccia & Levins, 2013) – the indirect interaction between bees and sheep, due to the grazing of sheep on forbs the bees feed on, or the indirect interactions between fish and plants via dragon-flies whose larvae are eaten by fish and whose adults prey on plant pollinators (Knight *et al.*, 2005). Indirect interactions thus generate intricate interconnections across communities, and could make them behave differently than if they were independent collection of populations or basic interaction motifs (Loreau, 2020; Liautaud *et al.*, 2019).

The interest in indirect interactions is not new (Bender *et al.*, 1984; Schmitz & Suttle, 2001). Their importance for community structure (Menge, 1995; Abrams *et al.*, 1996; Simmons *et al.*, 2019) and response to perturbations (Yodzis, 2000; Montoya *et al.*, 2009; Pires *et al.*, 2020) has been demonstrated on many occasions. Yodzis (1988) showed that the ability to predict the long-term ecological impact of a single-species perturbation (a cull on some fish population for instance) is tightly linked with the strength, and length, of indirect interactions present in the ecosystem. Knight *et al.* (2005) showed how indirect species interactions can couple terrestrial and aquatic ecosystems, while more recently, indirect interactions have been proposed as determinants of co-evolution in mutualistic networks (Guimarães *et al.*, 2017) and the diversity of natural communities (Bairey *et al.*, 2016).

Here, we take a step back to revisit and formalize the notion of indirect interactions. We compare direct interactions –the direct effect of species abundances on the growth rate of others, to net interactions – the long-term net impact that a change in demographic parameters of a population has on the abundance of another. We explain how indirect interactions connect these two notions by showing when and how a net interaction between two species integrates all indirect interaction pathways between them. But more importantly, formalizing these intuitive ideas leads us to a novel measure of collectivity that embraces the somewhat disconnected existing approaches to community complexity, organizational integration and indirect interactions.

Concretely, from a basic spectral feature of the matrix of pair-wise species interactions, we propose a measure of the general importance, in a given community, of indirect interactions. This measure determines the length of indirect pathways that contribute to community dynamics and patterns, a length that we call "the interaction horizon" and that mirrors the "environ" concept proposed by Patten (1982) from the era of theoretical ecosystem ecology.

We will illustrate, using simulations of community models, that our notion of collectivity explains the occurrence of intuitive signatures of collective community behavior, such as the depth of a perturbation's reach, the degree of temporal unpredictability or the contribution of biotic processes to realized species niches.

Our work revisits classic theoretical notions such as indirect interactions (Bender *et al.*, 1984; Menge, 1995), press perturbations, loop analysis (Puccia & Levins, 2013), complexity and stability (May, 1973). It clarifies when reductionist perspectives, focusing on particular species and small interaction motifs can, at least in principle, reliably scale-up to the community-level, or on the contrary, when there are fundamental obstacles facing such approaches (Bergelson *et al.*, 2021; Orr *et al.*, 2021). Our aim is to shed new light and inspire future work on the determination of the dynamical nature of ecological communities, or other complex systems, and their degree of organizational integration.

Collectivity and the interaction horizon

In this section we provide a heuristic introduction to a collectivity parameter, skipping technical details and focusing on intuitions. The reader interested in a direct derivation from general dynamical community models can skip this section and instead focus on the content of Box 1.

Let us encode the interaction network associated with a community as a matrix $A = (A_{ij})$, with $A_{ii} \equiv 0$. A_{ij} is a non-dimensional number that quantifies the direct interaction of species j on species i. It is crucial to see A_{ij} as a relative interaction strength: the ratio of inter- over intra-specific interactions. We emphasise this seemingly technical detail because it is key to properly define a notion of indirect interactions (see Box 1). In fact, if interactions had units, indirect interactions of different orders would themselves have different units, making them incomparable.

We define an indirect interaction of second order between species j and i through a third species k as $A_{ik} \times A_{kj}$, the product of the direct interaction of species j on species k with the direct interaction of species k on species i. More generally, following (Puccia & Levins, 1991) an interaction pathway of length n between species i and j is $A_{ik_{n-1}} \times ... \times A_{k_2k_1} \times A_{k_1j}$, where the intermediate species $k_1, ..., k_n$ need not all be different (there could be loops). Importantly, the sum over all such interaction pathways coincides with the element of the matrix A^n:

$$(A^n)_{ij} = \sum_{k_1,...,k_{n-1}} A_{ik_{n-1}} \times ... \times A_{k_2k_1} \times A_{k_1j} \tag{1}$$

We may note that if all the numbers A_{ij} are strictly smaller than one, the magnitude of indirect interactions will decay exponentially as their order n grows. But on the other hand, if the interaction network is sufficiently connected, the number of interaction pathways between species i and j (the number of terms in the sum shown in eq. 1) could also increase exponentially. It is therefore not clear if the sum of all indirect interactions will necessarily vanish, even if direct interactions are individually weak.

The concept of an indirect interaction is intuitive. But to understand why it precisely manifests in community models as Eq. (1), it is enlightening to consider the specific example of the equation verified by the steady state of a Lotka-Volterra system (see Box 1 for the general case):

$$N_i = K_i + \sum_{j=1}^{S} A_{ij}N_j;\ i = 1, ..., S \tag{2}$$

Here K_i denotes the carrying capacity of species i, encoding the environmental conditions as perceived by that species, in the absence of the other species from the community. If we introduce the vectors $K = (K_i)$ and $N = (N_i)$, Eq. (2) can be written in compact form as

$$N = K + AN \Leftrightarrow N = (\mathbb{I} - A)^{-1}K \tag{3}$$

Thus the carrying capacities of all species intertwine via the the matrix $(\mathbb{I} - A)^{-1}$ to determine the actual species abundances in the community context. For instance, a favourable environment (a large K_i) will not imply a large abundance if the environment is also favourable to a competitor. The matrix $(\mathbb{I} - A)^{-1}$ encodes all such effects, that is, all *net interactions* between species. On the other hand, if we had instead repeatedly applied Eq. (2) on itself we would have written a series highlighting the contribution of indirect interactions pathways, as defined in Eq. (1):

$$N_i = K_i + \underbrace{\sum_j A_{ij} K_j}_{\text{order } 1} + \underbrace{\sum_{j,k} A_{ik} \times A_{kj} K_j}_{\text{order } 2} + ... + \underbrace{\sum_{j,k_1,...k_{n-1}} A_{ik_{n-1}} \times ... \times A_{k_1 j} K_j}_{\text{order } n} + ...$$

Since this last expression should be equivalent to Eq. (3), and this for all K, we arrive at a classic matrix identity, known as Neumann's series (Reed *et al.*, 1972),

$$\mathbb{I} + A + A^2 + ... + A^n + ... = (\mathbb{I} - A)^{-1} \tag{4}$$

This series converges only under some specific conditions. When it does not converge, this means that we cannot meaningfully decompose the net interaction between two species (their effective interaction in the community context) as a sum of indirect interaction pathways.

The criteria for convergence gives us both a measure of the importance of indirect interactions, and a *definition* of collective integration. To make this idea more concrete, we first need to measure the magnitude of the various terms of the series, representing the overall strength of indirect interactions of all orders. This amounts to defining a matrix norm $||A^n||$. We then want to understand how this norm changes with the order n. Consider

$$\phi = \lim_{n \to \infty} ||A^n||^{1/n}, \tag{5}$$

that is, the rate of growth of the norm $||A^n||$ as n grows. If $\phi < 1$, as n grows, the overall contribution of indirect pathways will *eventually* decrease exponentially as ϕ^n. If $\phi > 1$, the sum over arbitrarily long pathways can be arbitrarily large (cf. Fig. 1).

Remarkably, ϕ does not depend on the particular choice of matrix norm. It is an intrinsic feature of the interaction matrix A: its spectral radius, the largest absolute value of its eigenvalues (Trefethen & Embree, 2020).

Here we propose an ecological interpretation of the spectral radius ϕ of a given interaction matrix. We call ϕ the *collectivity parameter*, because it determines the *interaction horizon* of species: the maximal length of interaction pathways that contribute to their net interactions(see Fig. 1). For systems for which $\phi > 1$, the interaction horizon is infinite, signaling the breakdown of the reductionist method of decomposing net effects into indirect interaction pathways, which we see as a reflection of the highly collective integration of such communities.

Signatures of collective integration

We now introduce four different, intuitive signatures of collective integration, which could conceivably be observed empirically. Not all four would be indicative of collectivity in a given system, but taken together they apply to a broad spectrum of ecological scenarios. All are driven by the collectivity parameter ϕ.

We use a large ensemble of Lotka-Volterra model communities, where community assembly starts from a random species pool, and leads to a steady state of coexisting species.

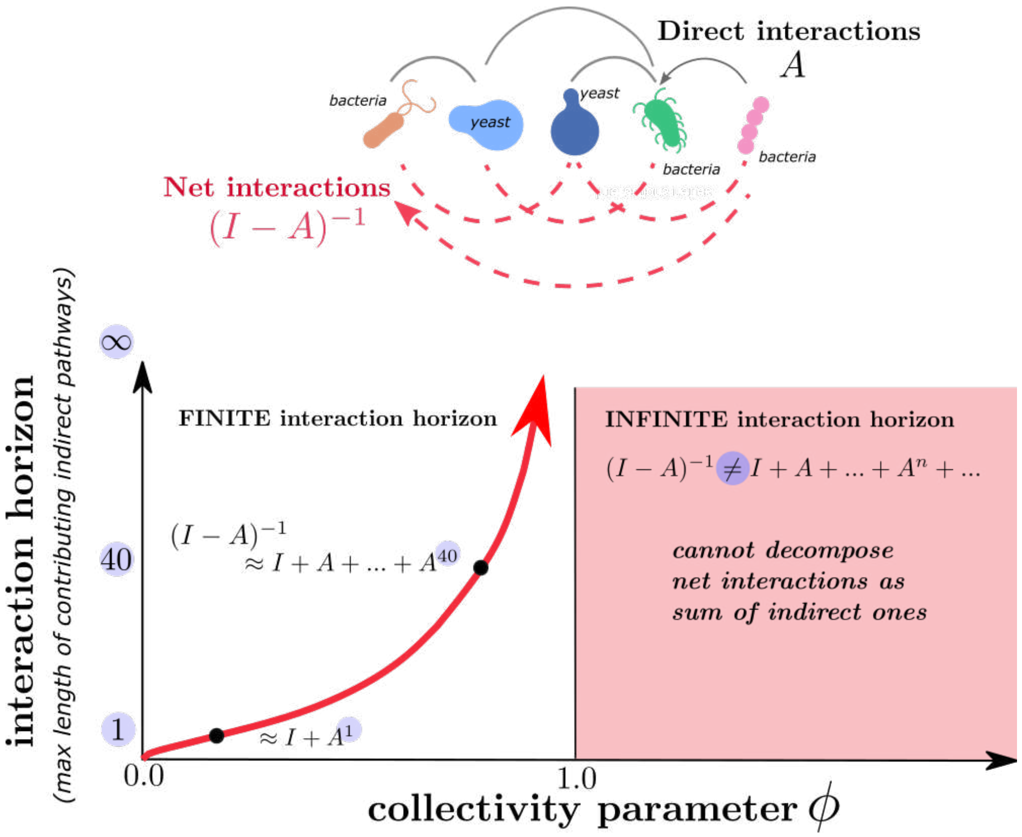


Figure 1: **The interaction horizon** is the maximal length of indirect interactions pathways that substantially contribute to net interactions (illustrated here in a hypothetical yeast-bacteria community). The horizon is directly determined by the collective parameter as $\log \epsilon / log \phi$, where $\epsilon < 1$ is an arbitrary threshold value. The horizon gives the lowest order of interactions for which the maximal contribution is negligible (that is, smaller than ϵ), and it diverges as ϕ approaches 1. Beyond this point it no longer makes sense to decompose net interactions as a sum of indirect pathways.

Conveniently, in Lotka-Volterra models the interaction matrix A is a set of parameters. So by prescribing an ensemble of matrices to pick from, we can easily generate systems along a gradient of collective integration.

We will consider a gradient of interaction strength (and heterogeneity), with 50 different values of overall interaction strength along the gradient, each with 100 random communities, making up 5000 communities in total. Each community starts with $S = 50$ species, and we set 80% of the interactions to zero, so that we have a sparse interaction matrix. To specify the interactions between connected species, we follow the tradition of Random Matrix Theory applied to ecology (May, 1972; Allesina & Tang, 2012; Bunin, 2017; Barbier *et al.*, 2018). Let y be the control parameter that we vary to create a gradient of interaction strength, with values ranging, in uniform intervals, between 0.02 and 1. We define three parameters of random interactions: $\text{std}(A_{ij}) = y/\sqrt{S}$; $\text{mean}(A_{ij}) = -y/S$; and $\text{corr}(A_{ij}, A_{ji}) = -1$. This gives us asymmetrical interactions between species, that are increasingly negative and at the same time more varied. With this protocol we generate communities with a collectivity parameter ϕ ranging between 0 and 2, thus allowing us to showcase different, generic, aspects of low and high collective integration in communities.

Increase in effective connectance

If a community is collectively integrated, substantial net interactions should connect any two species, even if these species do not directly interact. In other words, the effective connectance of the community should be much larger than the one of the direct interaction network. To illustrate this idea we define effective connectance as the Shannon diversity of net interactions (the off-diagonal elements of the matrix $(\mathbb{I} - A)^{-1}$) relative to the one of direct interactions (the matrix A). Values larger than one indicate an increase in connectance when considering net interactions instead of direct ones. We see in Figure 2 that the collective parameter ϕ is indeed closely related with the factor that relates effective and direct connectance (left panel). As collectivity grows, net interactions become more and more uniformly distributed leading to a larger and larger effective number of connections between species (Fig. 2 right panel).

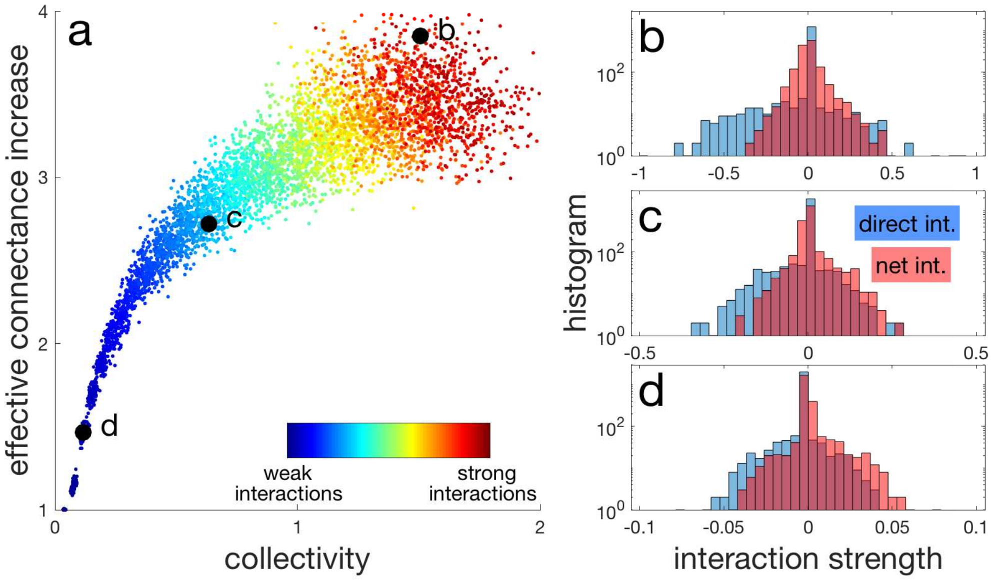


Figure 2: **Growth in the effective number of net interactions, relative to the number of direct ones, as collectivity increases**. Left panel shows the effective connectance – the ratio between the Shannon diversity index of the direct interactions and net interactions, as defined in the main text. Black circles highlight several communities that are considered in the right panels. The right panels show the histograms of direct and net interactions, overlayed, for these communities. Note the logarithmic scale on the y-axis and the changes across panels of the x-axis.

Perturbation depth

Collective integration means that species are interdependent. As a consequence, a perturbation targeted on a given species will likely propagate deep into the interaction network (Bender *et al.*, 1984). Experimentally one could remove a species, and monitor the long term response of others, as a functions of their interaction distance $d(i, j)$ from species removed ($d(i, j) = 1$ if j interacts with i, $d(i, j) = 2$ if i and j are indirectly connected via a third species, and so on). Denoting $N_{j\setminus i}$ the long term abundance of species j after the removal of species i, this leads us to the notion of *perturbation depth*

$$\mathrm{PD} = \frac{\sum_j d(i,j)\left|N^*_{j\setminus i} - N^*_j\right|}{\sum_j \left|N^*_{j\setminus i} - N^*_j\right|} \tag{6}$$

which we can average over all species removal experiments in that community. Perturbation depth equals 1 when only directly interacting species are affected by the removal, and has higher values the more species further away in the interaction network are heavily impacted.

In Fig. 3 (left panel) we demonstrate a good agreement between this observable signature of collective integration and the collectivity parameter ϕ. As collectivity grows, the brunt of the perturbation effect is shared with more distant species, and not only supported by those directly in contact with the removed node (right panel). An obvious caveat of perturbation depth is that it only applies to sufficiently sparse networks –if all species are connected, this notion is useless.

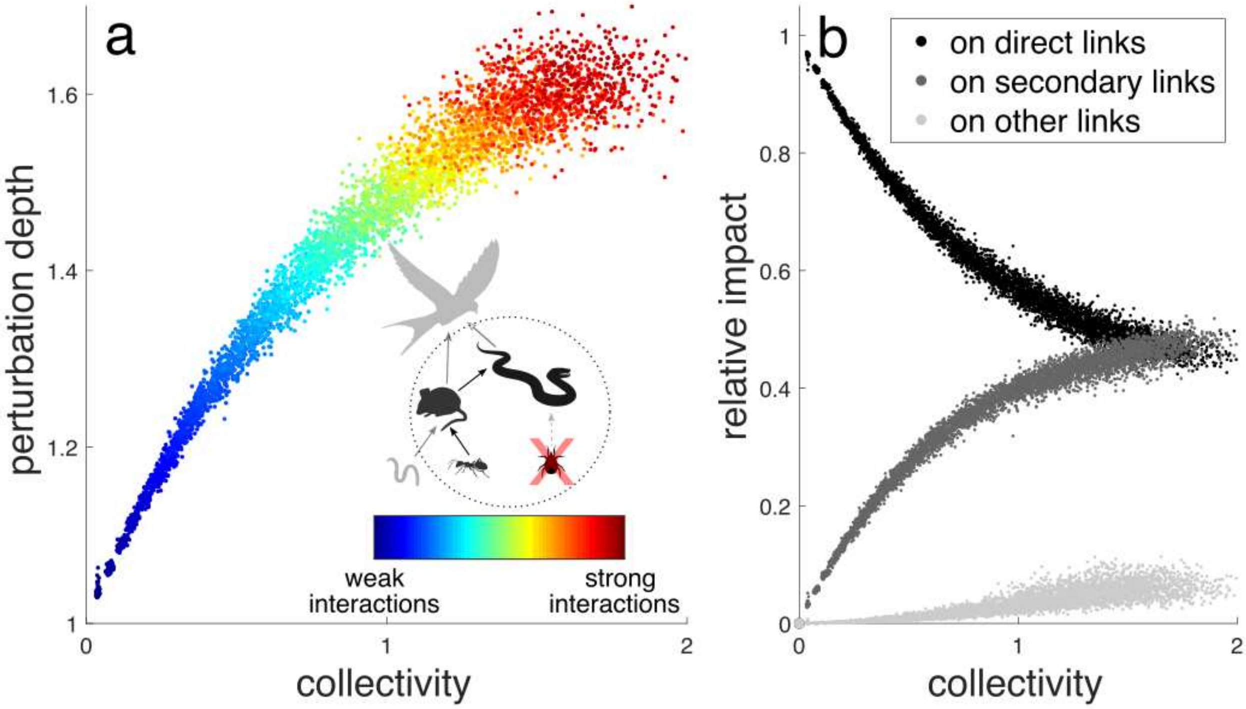


Figure 3: **Perturbation depth and collectivity**. For various communities, the effect of removing a single species is seen across the community. Left panel shows the perturbation depth, a measure of how deep into the network of species interactions does the perturbation reach. Right panel shows the average effect on the species in the community (all except the one species removed), partitioned into three groups: black for species directly interacting with removed species, dark gray for species directly interacting with the species in the black group, light gray for all other species. As collectivity increases the average effect on a given species becomes equal, regardless of its grouping (i.e. its position in the community structure), and therefore the perturbation depth increases (i.e. the effect of the perturbation if felt throughout the community). See Appendix for more details.

Temporal unpredictability

Indirect interactions between species require some time to take effect. Thus, collective integration is expected to leave a signature in the relationship between short- and long-term response to a perturbation. Consider a persistent change in abiotic conditions. At the moment the perturbation is applied, a given species' population will react to the induced change in its intrinsic growth rate. At slightly longer time scales, direct interactions between species will take effect. With time, longer and longer interaction pathways can start to play a role in a population's response. However, if the strenght of indirect intercations rapidly decays with their lenght, the latter will not substantially change the course of population dynamics; the long term outcome could have been extrapolated from the short term response. In this view, the more collectively integrated the community, the less predictable the long-term response of a species should be. This leads us to a notion of *temporal unpredictability*, which quantifies the decorrelation, across species in a perturbed community, between long-term predictions based on short-term responses, and the actual long-term dynamics of the community.

We test this idea by considering a community at equilibrium in which a perturbation changes the intrinsic growth rates of all species randomly. We then measure the correlation between a vector of short-term response extrapolation R_S and a vector of the actual long-term response R_L (see Appendix), and define temporal unpredictability as the complement of that correlation:

$$\mathrm{TU} = 1 - \mathrm{corr}(R_S, R_L) \tag{7}$$

In Fig. 4 we confront this notion to our collectivity parameter ϕ. We see that the two are closely related, with unpredictability increasing steadily as collectivity grows, reflecting the fact that trajectories can, as indirect interactions come into play, change tendencies through time (right panel). As always there is a caveat. If direct interactions are mediated by slow latent variables, such as unobserved species or modified environmental variables, time and length of interaction pathways need not be related. Collectivity thus leaves a univocal signature in temporal trends only if a separation of time scales exists between the factors that mediate direct interactions, and the population dynamics.

Biotic contribution to the realized niche

If species do not interact, only the abiotic environment (i.e. what cannot be attributed to the rest of the community) determines the growth and abundance of a species. In general however, species change the environmental conditions perceived by other species. Intuitively, we can expect that the stronger the collective integration of the community, the more important and intricate this biotic contribution becomes (Levine *et al.*, 2017). As a final illustration of observable consequences of collective integration, we propose a simple formalization of this expectation. In our simulations, we first quantify the extent to which the abundance of a species N_i, expressing its realized niche, is explained by its carrying capacity K_i, expressing the fundamental niche, i.e. the environmental conditions perceived by this species, in the absence of the other species from the considered community. In an experimental setting this would amount to comparing mono-cultures to polycultures (second column of Fig. 5).

Conversely, we can quantify the extent to which the biotic environment alone (i.e. species interactions) explains species abundance. To do so, we compare N_i to the abundance that it would have if all species perceived the exact same environmental conditions (third column of

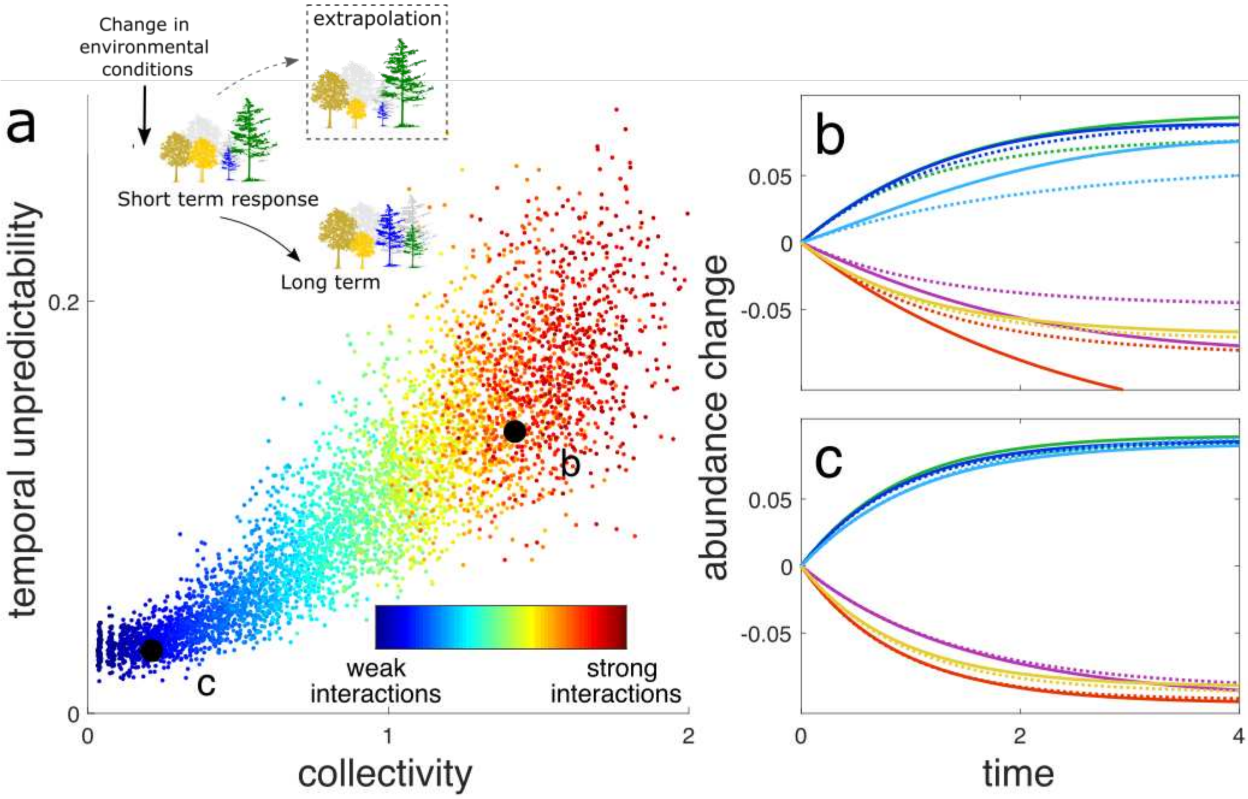


Figure 4: **Temporal unpredictability between short-term and long-term response to perturbation**. For various communities, the ability to predict the long-term response of a community to a perturbation from its short-term response is evaluated and shown. Left panel shows the temporal unpredictability, which gives a score of 0 for a perfect correspondence between short-term and long-term response. Black circles show the collectivity and temporal unpredictability for two communities, with the corresponding dynamics shown in the right panels. Right panels show the change in abundance for 6 species in each community, where the dashed lines show the extrapolated dynamics based on the short term fit (using the first 0 to 0.5 time units), whereas the solid lines show the actual dynamics. With higher collectivity the long-term behavior becomes less predictable, see Appendix for more details.

Fig. 5). This amounts to asking whether a species abundance is explained by its *centrality* within the interaction network (Sharkey, 2017).

To synthesize the two above perspectives, we start from the relative yield of a given species, $\eta_i = N_i/K_i$. The absolute difference between relative yield and unity is a measure of the net effect, on species i, of the biotic environment. We measure this effect integrated over the community by summing this difference over all species. To make the result comparable across communities we normalize by the sum of relative yields. This gives a quantification of the Biotic Contribution to species realized niches:

$$\mathrm{BC} = \frac{\sum_i (\eta_i - 1)^2}{\sum_i \eta_i^2} \tag{8}$$

which is similar to the relative Euclidean distance $||N - K||/||N||$ between the realized community state N and what it would have been without interactions, K.

In the first panel of Fig. 5, we see that the collectivity parameter and the strength of the biotic niche, defined in Eq. 8, closely follow one another. Furthermore, we see that for communities with collectivity parameter close or larger than 1, species abundances are not at all explained by the abiotic environment that they perceive, but instead they are almost entirely controlled by the biotic environment set by the whole community. The major caveat here is the requirement of a notion of carrying capacity, which makes senses for, say, plants but is ill-defined when considering consumer species.

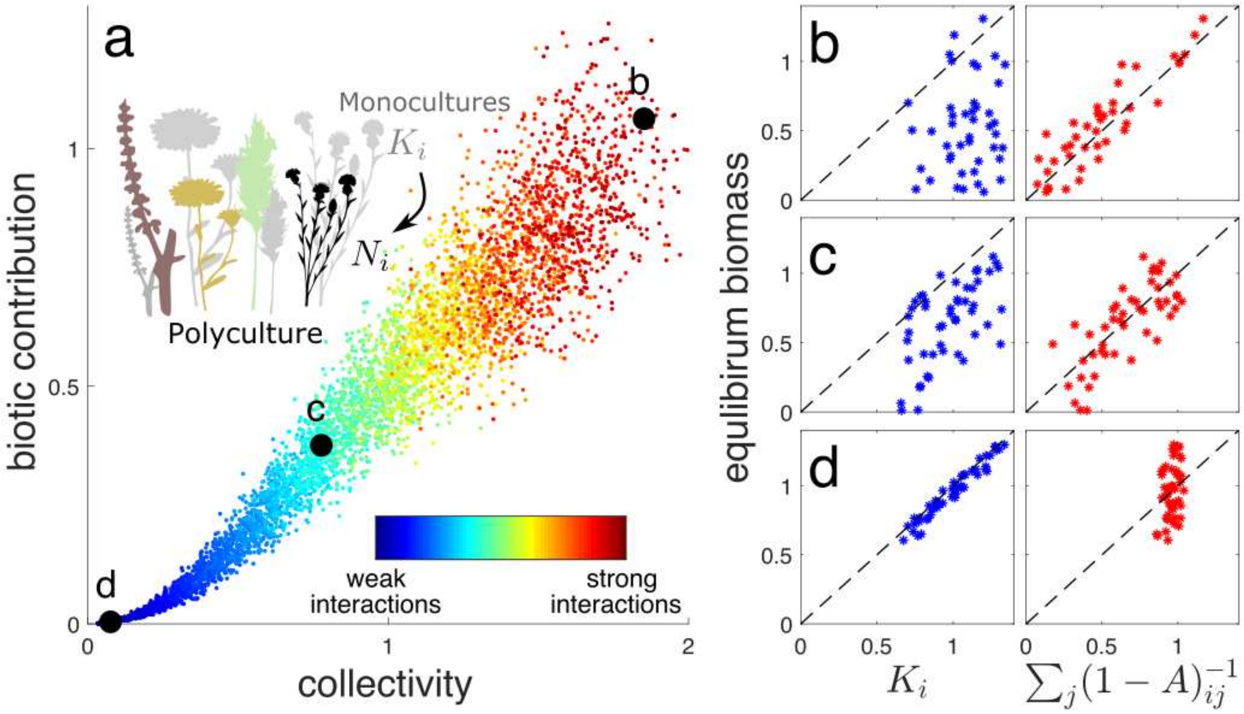


Figure 5: **Biotic Contribution to species realized niches.**The determinants of the community's species abundance at equilibrium are evaluated. (a) the biotic contribution to the realized niche (eq. 8), with black circles highlighting several communities that are considered in the right panels. (b)-(d) Species equilibrium abundance for different communities (corresponding to black circles in panel a), compared with its carrying capacity (left in blue), or by contrast, with its abundance if all species had the same carrying capacities, so that differences in abundances are caused by species interactions only (on the right in red). Dashed black line shows the 1:1 line. See Appendix for more details.

Collectivity, complexity and stability

If collectivity determines the contributions of indirect interactions in shaping long-term population dynamics, what are the factors that drive collectivity? Clearly, the connectivity of the interaction network as well as the strength of interactions must be determining factors. This is indeed the case and by making this intuition more precise we can relate collectivity to the *complexity* notion of May (1973) and further clarify its relatedness and difference with a central notion of theoretical ecology: *stability.*

From its definition Eq. (5) (or Eq. 20 in Box 1), ϕ is essentially[1] the norm of the interaction matrix A. In fact, it always holds that $\phi \leq ||A||$, where the norm in question is the spectral-norm, still an abstract quantity, defined as the maximal amplification of vectors by that matrix

---

[1] Exactly for normal matrices, such as symmetrical ($A_{ij} = A_{ji}$) or anti-symmetrical ($A_{ij} = -A_{ji}$) matrices.

(Reed *et al.*, 1972). However, we can relate the spectral-norm to a much simpler notion via the following equivalence relation, valid for any matrix (Reed *et al.*, 1972):

$$\sqrt{1/r}||A||_F \leq ||A|| \leq ||A||_F \tag{9}$$

where r is the rank of the matrix, which for us will always be the number of coexisting species S, and $||A||_F$ is the Frobenius norm of A: the sum of its squared elements. Then, if C is the connectance of the network and ξ^2 the second moment of interactions between connected species, we get that

$$\sqrt{1/S}||A||_F = \sqrt{C(S-1)\xi^2} \equiv \mathcal{C},$$

which is the measure of complexity, that we note $\mathcal{C}$, proposed by May (1972). From this observation, and using the fact that $A_{ii} \equiv 0$, we can transform the general equivalence relationship Eq. (9) to deduce a similar relationship between complexity and collectivity (see appendix):

$$\mathcal{C} \lesssim \phi \leq \mathcal{C}\sqrt{S-1} \tag{10}$$

The upper bound is sharp while the lower bound is an approximate one (i.e. sharp and impossible to cross only for normal matrices, such as symmetric matrices). The upper bound states that May's complexity, once scaled by $\sqrt{S-1}$, sets the collectivity potential that the structure of the interaction network can achieve. Distributing all interactions evenly across species saturates the upper bound, thus maximizing collectivity[2]. The interpretation of the approximate lower bound is a bit more subtle. If the interaction matrix is perfectly symmetrical $A_{ij} = A_{ji}$ (or perfectly anti-symmetrical $A_{ij} = -A_{ji}$), then $\phi = ||A||$ and it follows that the bound is sharp, setting the minimal collectivity that can be achieved while maintaining complexity. But in general, this lower bound can be breached and should rather be be seen as the collectivity that 'comes for free' if the matrix was fully random. Interaction structure can can reduce collectivty far bellow the random baseline. A simple example is that of a triangular (i.e. perfectly nested) interaction structure (as in the competition-colonization trade-off model of Tilman (1994)). One species affects all others but is affected in return by no one, and so on until the last species that is affected by all but that affects no one. In this case, due to the impossibility of feedbacks up the species hierarchy, the interaction horizon remains finite no matter how strong or varied the direct pair-wise interactions are.

But, if ϕ is associated with May's complexity measure, should it not be directly related to stability, at least sensu May (1972)? Are we not rediscovering the same notion and giving it a different name? The linear stability criterion is that all eigenvalues of the Jacobian matrix at the steady-state must have a negative real part (Lyapunov, 1892). This is essentially equivalent to all eigenvalues of A having a *real part* smaller than 1. Collectivity ϕ is the largest eigenvalue *modulus*. As is made clear once represented graphically (see middle inset of Fig. 6), if the system is unstable ϕ is necessarily larger than 1. However, even if ϕ is large,

---

[2] Consider a model, where all species interact with equal strength $-a < 0$. The eigenvalues of this interaction matrix are $-(S-1)a$ and a. The stability criterion here becomes that $a < 1$. If there are only two species, instability is reached when $a = 1$ (mutual exclusion) and corresponds to a $\phi = 1$. Collectivity and instability go hand in hand. For larger S the stability criterion remains that $a < 1$ but now $\phi = (S-1)a$. Thus, the more species in the community the more collectively integrated it becomes, and can exceed the critical value of 1, while remaining stable. Here $||A||_F = \sqrt{S(S-1)}a = \sqrt{\frac{S}{S-1}} \times \phi = \frac{\phi}{\sqrt{1-1/S}}$. This shows that this example saturates the upper bound in Eq. (10), demonstrating that it is indeed sharp.

the real part of the associated eigenvalue can still be arbitrarily small. So instability implies a high degree of collective integration but the converse is not true. Large collective integration does not imply instability (right panel of Fig. 6).

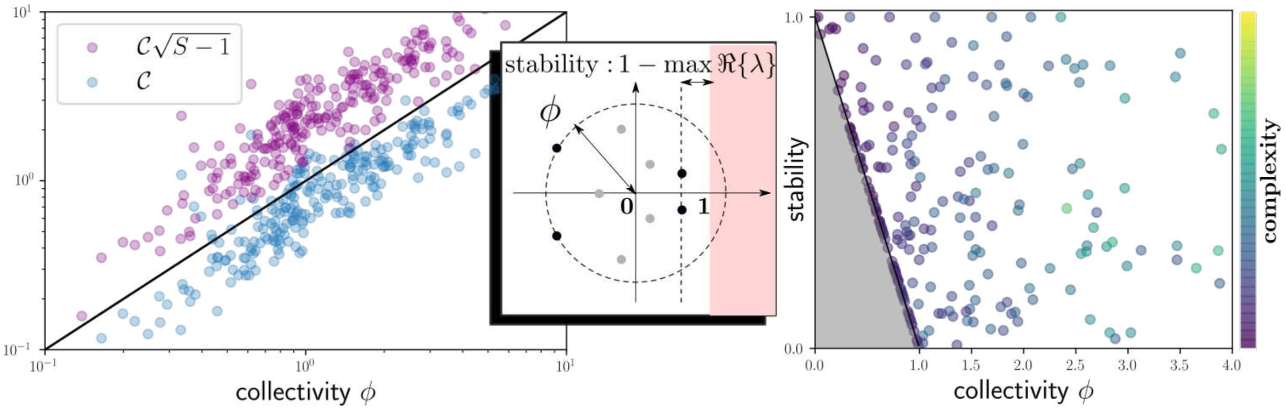


Figure 6: **Left: May's complexity $\mathcal{C}$ sets the bounds of collectivity**, illustrated here for 242 random stable matrices (out of 1000 generated) of size 5×5 to 15×15, representing interaction matrices A of various connectance, interaction strength, variance, and pair-wise symmetry (notice the log-log scales). In purple the sharp upper bound $\mathcal{C}\sqrt{S-1}$ where S is the size of the community (purple points are always above the one:one line in black). In blue the approximate lower bound $\mathcal{C}$ (blue points mostly remain below the one:one line). **Middle inset:** Stability is here a distance to instability $1 - \max \Re\{\lambda\}$, where $\max \Re\{\lambda\}$ is the maximal real part over all eigenvalues $\lambda \in \mathbb{C}$ of the interaction matrix. If this real part attains 1, the community is unstable –shaded region is the instability domain. Collectivity is instead the radius of the smallest disc centered on 0 that contains all eigenvalues. **Right: collectivity and (in)stability are not equivalent.** Same matrices as on the left panel. We see that large values of collectivity are allowed even if we restrict to stable systems. The edge of the grey region represents $y = 1 - x$ which is what would be expected if collectivity and stability where associated to the same eigenvalue.

Discussion

Species are embedded in complex networks of interactions, but this does not tell us how truly interdependent they are, and on what time scales. To understand the dynamics of a species, is it enough to account for the dynamics of its direct interaction partners (prey, predators or direct competitors)? The answer to this question lies in the analysis of indirect interactions, and we proposed that this analysis amounts to measuring the degree of *collective integration* of ecological communities.

In nature, the inter-dependency between species occurs not only though direct interactions, via predation, facilitation or competition, but through potentially much longer indirect interaction pathways. The question is not whether long indirect pathways exist. Most of the time, an indirect interaction pathway can connect any two species. The question is whether or not those indirect pathways significantly contribute, over relevant time-scales, to population dynamics. If they do matter, the effective interaction between two species indirectly involves many others: the state of a species is dependent on many (if not all) other species in the community. Since interaction pathways are long, perturbations may also propagate further: a perturbation targeted on a single species can have far-reaching consequences in the community. In such collectively integrated communities the abundances of species will be only faintly related to what they would have been if grown in isolation; in other words, biotic effects (by opposition to the abiotic environment) could play a predominant role in shaping abundance distributions patterns.

Here, we quantified the dynamical importance of indirect interaction pathways as the spectral radius ϕ of the direct interaction matrix. This synthetic feature of the interaction network determines the rate of growth of the contribution of indirect interactions as their length (or order) increases. If ϕ is smaller than one, long indirect interactions pathways are negligible. By contrast, if ϕ is larger than one, arbitrary long interaction pathways can have as much importance as shorter ones. This signals the breakdown of the reductionist method of decomposing net interactions into a few dominant direct or indirect interaction pathways.

This measure of collective integration ϕ drives empirically relevant phenomena. Using a broad set of simulation models we showed that the collectivity parameter ϕ is strongly correlated with *Perturbation depth*, defined as the network distance covered by perturbations initially affecting only a single species. We also showed that ϕ increases *Temporal unpredictability*, defined as the discrepancy between species long-term behaviour after a change in environmental conditions, and the extrapolation of their dynamics based on their short-term response to the environmental change. We finally related the value of the collectivity parameter with a measure of *Biotic Contribution* to species realized niches, defined to quantify the degree to which abundances of species grown in monocultures predict species abundances when grown in the presence of other species.

There is a direct connection between complexity sensu May (1972), which is essentially a measure of total, absolute, interaction strength in the system, and collectivity. Complexity sets the collectivity potential that network structure can realize. In this sense, collectivity encapsulates two intertwined notions: random and structured complexity. The former sets the baseline from which the latter will depart: we identified structures that push towards maximal collectivity: evenly distributing interactions do just that, as well as increasing pairwise symmetry. By contrast, highly nested interactions can reduce collectivity drastically, and even completely (in the case of a triangular interaction matrix (Tilman, 1994)). These examples suggest a novel perspective on the impact of network structure for dynamical patterns, that could help move beyond the long-lasting complexity-stability debate (May, 1972) by instead asking whether ecological network structures tend to make communities more or less collectively integrated than if they had been randomly assembled.

In network studies, there is a growing interest in higher order interactions, which generalize pair-wise interactions to include more than two interacting partners at once. This notion challenges the relevance of viewing communities as a graph, with nodes connected by directed edges (Battiston *et al.*, 2021), since now the existence and/or weight of an edge between two nodes is conditioned on the status of other nodes. For example, planktivorous fish i might prey on zooplankton j if predatory fish k is not present, but the presence of k prevents i from preying on j. Our focus here was indirect interactions, and despite a very similar terminology, they are not equivalent to higher-order interactions. The latter occur when considering non linear extensions of Lotka-Volterra models, so that the coupling terms between dynamical variables depend on the product of several variables at once. Studying higher order interactions amounts to understanding the effects of such non-linearities. For us, this implies that the interaction matrix, as defined in Box 1 could be state-dependent, and not just dependent on the species composition (as is the case for Lotka Volterra models). As is made clear in the formal derivation proposed in Box 1, this does not affect the definition of collective integration. From this perspective, our work may offer a novel direction of investigation: understand how higher order interactions affect the collectivity of steady sates.

In collectively integrated communities, long indirect pathways between species, whose importance unfolds over variable time scales, do not favour predictive power. As has been shown before (Yodzis, 1988, 2000; Barabás *et al.*, 2014), this is certainly true if one wishes to accurately predict the effect that a perturbation will have, in the long-term, on a given species when the latter is embedded in a complex ecosystem. But this typical sensitivity of individual variables need not imply that all properties of a community or ecosystem are sensitive and/or unpredictable (Daugaard *et al.*, 2022). We know, in particular, that many aggregate features of complex models are robust to uncertainty in parameters (Barbier *et al.*, 2018). From our work, we can predict when those long indirect interaction pathways cannot be ignored or even simply added up. In such instances there is no choice but to change perspective, and focus our efforts towards robust ecosystem or community-level properties of the natural system under study (Goldford *et al.*, 2018; Bergelson *et al.*, 2021; Sanchez *et al.*, 2022).

Perspectives

Our measure of collective integration is defined on isolated community models, where species populations are dynamical variables that deterministically change through time. This is a drastic simplification of real communities, that are open to migration, structured spatially, and neither deterministic nor stable (Hastings, 2004).

It is not obvious how to translate our mathematical analysis to transient, far-from-equilibrium dynamics or draw conclusions on the role played by spatial structure – and spatial-scale – in determining the degree of collective integration of ecological systems. For the latter, this connects with the research program surrounding the meta-community concept (Holyoak *et al.*, 2005) and the quest to better understand and theorise the spatial scaling of ecological interactions (Gravel *et al.*, 2016; Galiana *et al.*, 2018, 2022). This is a promising direction, that could help better formulate the scale transitions of ecological patterns, as it is commonly thought that at large scales, ecological systems are mostly determined by abiotic drivers, such as long-term climate patterns, thus suggesting (if only vaguely) a negative collectivity-scale relationship.

Extending our theory of collective integration to transient states is certainly challenging. Powerful dynamical theories do exist to tackle transients in complex systems (Roy *et al.*, 2019). But even without abandoning the notions of stationarity or equilibria, we could expand our formalism to address the time-scale dependency of collectivity. A way forward may be the study of power-spectra of ecological time-series: decomposing the temporal fluctuations of populations over various times scales to infer the interaction structures that generate such signals (Krumbeck *et al.*, 2021). Here we suspect that variations over longer time scale reflect the collective nature of communities more than those at much shorter time scales, as the latter would not allow for indirect interactions to manifest, as already illustrated by the results shown in Fig. 4.

Even in stationary states, and without spatial structure, there are other forms of interesting collective behaviour that we did not consider. Communities could be structured by interactions between mutually excluding species (Fried *et al.*, 2016) which may lead to rich patterns (the ghost of competition past) and highly non trivial collective behaviour (Bunin, 2021).

Nonetheless, the idea that collective dynamical behaviour should manifest through indirect interactions, and that a high degree of collective integration implies the impossibility for causal links between variables to be decomposed in smaller parts, is a general one. Our contribution was to show how this idea can make way to a quantitative theory, at least in simplified settings.

Acknowledgements

ML and JFA were supported by the "Laboratoires d'Excellences (LABEX)" TULIP (ANR-10-LABX-41). NG received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement BIOFOODWEB (No 101025471). JFA thanks Mathew Leibold and members of the Leibold Lab for their encouraging and thoughtful comments during the writing of this manuscript.

Box 1: **direct, net and indirect interactions in stable community models**

Consider a model that specifies the growth rate g_i of all species, as a function of their joint abundances $\boldsymbol{N} = (N_i)$:

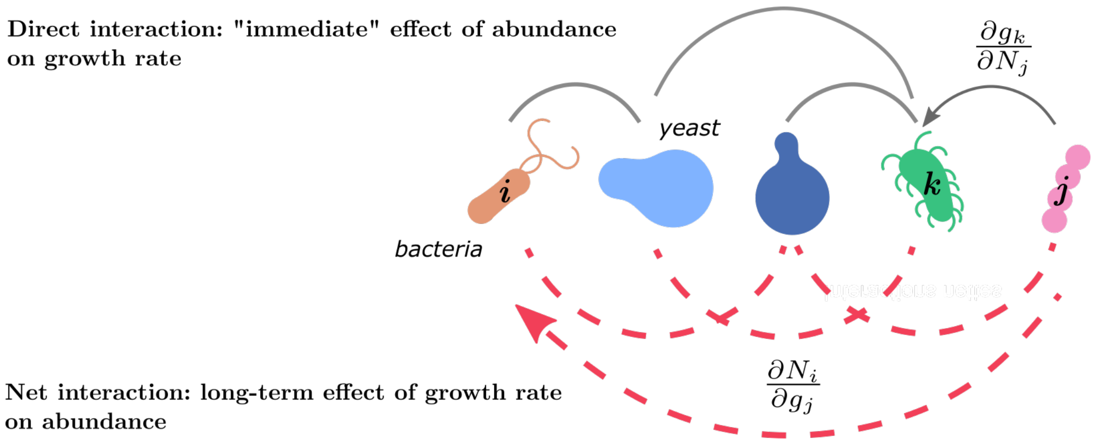
$$\frac{1}{N_i}\frac{dN_i}{dt} = g_i(\boldsymbol{N}) \tag{11}$$

In Lotka-Volterra (L-V) models $g_i(\boldsymbol{N}) = r_i + \sum_j a_{ij}N_j$, with $a = (a_{ij})$ representing per-capita interactions and $\boldsymbol{r} = (r_i)$ the vector of species intrinsic growth rates. We assume that the community is in a steady state $\boldsymbol{N}^*$, so that $g_i(\boldsymbol{N}^*) = 0$. In that state, we want to define direct and net species interactions, relate them to one-another, and show *in what sense net interactions emerge as a sum of indirect* ones, a claim that we use to quantify the collective integration of the community.

Direct interactions reflect the sensitivity of the growth rate of one species, to a change in abundance of another. In mathematical terms, this amounts to defining the matrix of partial derivatives

$$\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{N}} := \left(\frac{\partial g_i}{\partial N_j}\right) \tag{12}$$

which for L-V models, coincides with the matrix $a = (a_{ij})$.



Net interactions *are the reciprocal of direct interactions*: the long-term sensitivity of the abundance of a species, to a permanent shift in the growth rate of another (Novak *et al.*, 2016; Montoya *et al.*, 2009). In matrix form, net interactions can be written as

$$\frac{\partial \boldsymbol{N}^*}{\partial \boldsymbol{g}} := \left(\frac{\partial N_i^*}{\partial g_j}\right) \tag{13}$$

To clarify the meaning of Eq. (13), imagine applying a small press perturbation $\delta\boldsymbol{g}$ on the species' growth rates. In contrast with the way direct interactions are defined, we now *let community dynamics play out*, ultimately leading to a shift in equilibrium abundances $\delta\boldsymbol{N}^*$, so that

$$\boldsymbol{g}\left(\boldsymbol{N}^* + \delta\boldsymbol{N}^*\right) + \delta\boldsymbol{g} = 0 \Leftrightarrow \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{N}} \cdot \delta\boldsymbol{N}^* + \delta\boldsymbol{g} = 0 \tag{14}$$

This expression can be inverted to show that the matrix of net interactions is indeed the inverse of the matrix of direct interactions:

$$\frac{\partial \boldsymbol{N}^*}{\partial \boldsymbol{g}} = \left(-\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{N}}\right)^{-1} \tag{15}$$

In L-V models this matrix also determines the steady state $\boldsymbol{N}^* = (-a^{-1}) \cdot \boldsymbol{r}$.

Indirect interactions and the collectivity parameter. Direct and net interactions have reciprocal units. Furthermore, if we multiply the direct interaction between species i and j, with the direct interaction between species j and k this would change dimensions, and define an indirect interaction between species i and k that cannot be compared to neither direct nor net interactions. However, by defining direct interactions *relatively to self-regulation*, defined for any species i as $-\frac{\partial g_i}{\partial N_i}$, we can revisit the connection between direct and net interactions such that a coherent notion of indirect interaction emerges. Let direct, *non-dimensional* interactions be

$$A_{ij} := \frac{\partial g_i}{\partial N_j} / (-\frac{\partial g_i}{\partial N_i});\ i \neq j;\ (A_{ii} = 0) \tag{16}$$

In L-V models this corresponds to $A_{ij} := a_{ij}/(-a_{ii})$. If D is the diagonal matrix encoding species self-regulation, and $\mathbb{I}$ the identity matrix, direct interactions can be written as

$$\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{N}} = D \cdot \underbrace{(-\mathbb{I} + A)}_{\text{non-dimensional direct interactions}} \tag{17}$$

which indeed have the same units as D. From Eq.(15), it follows that

$$\frac{\partial \boldsymbol{N}^*}{\partial \boldsymbol{g}} = \underbrace{(\mathbb{I} - A)^{-1}}_{\text{non-dimensional net interactions}} \cdot D^{-1} \tag{18}$$

If the elements of A are small, the *non-dimensional* net interaction matrix $(\mathbb{I} - A)^{-1}$ can then be written as a convergent infinite series

$$(\mathbb{I} - A)^{-1} = \mathbb{I} + A + A^2 + ... \tag{19}$$

This series enables us to define **indirect interactions** of order k as the elements of A^k. Indeed $(A^k)_{ij}$ is the sum of all non-dimensional interaction pathways of length k that lead from species j to i (allowing for loops), drawn in the interaction network. It is in the precise sense of Eq.(19) that net interactions emerge as a sum of indirect ones. Our measure of collectivity ϕ is the spectral radius of A (Trefethen & Embree, 2020), namely:

$$\phi = \max\{|\lambda|;\ \lambda \in \text{spect}\,(A)\} \tag{20}$$

which controls the rate of convergence of the series Eq.(19), and thus **the order of indirect interactions that contribute to net interactions**.

References

Abrams, P.A., Menge, B.A., Mittelbach, G.G., Spiller, D.A. & Yodzis, P. (1996). The role of indirect effects in food webs. In: *Food webs*. Springer, pp. 371–395.

Agrawal, A.A. (2001). Phenotypic plasticity in the interactions and evolution of species. *Science*, 294, 321–326.

Allesina, S. & Tang, S. (2012). Stability criteria for complex ecosystems. *Nature*, 483, 205–208.

Bairey, E., Kelsic, E.D. & Kishony, R. (2016). High-order species interactions shape ecosystem diversity. *Nature communications*, 7, 1–7.

Barabás, G., Pásztor, L., Meszéna, G. & Ostling, A. (2014). Sensitivity analysis of coexistence in ecological communities: theory and application. *Ecology Letters*.

Barbier, M., Arnoldi, J.F., Bunin, G. & Loreau, M. (2018). Generic assembly patterns in large ecological communities. *Proceedings of the National Academy of Sciences*, 115, 2156–2161.

Battiston, F., Amico, E., Barrat, A., Bianconi, G., Ferraz de Arruda, G., Franceschiello, B., Iacopini, I., Kéfi, S., Latora, V., Moreno, Y. *et al.* (2021). The physics of higher-order interactions in complex systems. *Nature Physics*, 17, 1093–1098.

Bender, E.A., Case, T.J. & Gilpin, M.E. (1984). Perturbation experiments in community ecology: theory and practice. *Ecology*, 65, 1–13.

Bergelson, J., Kreitman, M., Petrov, D.A., Sanchez, A. & Tikhonov, M. (2021). Functional biology in its natural context: A search for emergent simplicity. *Elife*, 10, e67646.

Brown, J.H., Whitham, T.G., Ernest, S.M. & Gehring, C.A. (2001). Complex species interactions and the dynamics of ecological systems: long-term experiments. *Science*, 293, 643–650.

Bunin, G. (2017). Ecological communities with Lotka-Volterra dynamics. *Physical Review E*, 95, 042414.

Bunin, G. (2021). Directionality and community-level selection. *Oikos*, 130, 489–500.

Clements, F.E. (1916). *Plant succession: an analysis of the development of vegetation*. 242. Carnegie Institution of Washington.

Daugaard, U., Munch, S., Inauen, D., Pennekamp, F. & Petchey, O. (2022). Forecasting in the face of ecological complexity: number and strength of species interactions determines forecast skill in ecological communities.

Fried, Y., Kessler, D.A. & Shnerb, N.M. (2016). Communities as cliques. *Scientific reports*, 6, 1–8.

Galiana, N., Lurgi, M., Bastazini, V.A., Bosch, J., Cagnolo, L., Cazelles, K., Claramunt-López, B., Emer, C., Fortin, M.J., Grass, I. *et al.* (2022). Ecological network complexity scales with area. *Nature Ecology & Evolution*, 6, 307–314.

Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.J., Leroux, S., Cazelles, K., Gravel, D. & Montoya, J.M. (2018). The spatial scaling of species interaction networks. *Nature Ecology & Evolution*, 2, 782–790.

Gleason, H.A. (1926). The individualistic concept of the plant association. *Bulletin of the Torrey botanical club*, pp. 7–26.

Goldford, J.E., Lu, N., Bajić, D., Estrela, S., Tikhonov, M., Sanchez-Gorostiaga, A., Segrè, D., Mehta, P. & Sanchez, A. (2018). Emergent simplicity in microbial community assembly. *Science*, 361, 469–474.

Gravel, D., Massol, F. & Leibold, M.A. (2016). Stability and complexity in model meta-ecosystems. *Nature communications*, 7, 1–8.

Guimarães, P.R., Pires, M.M., Jordano, P., Bascompte, J. & Thompson, J.N. (2017). Indirect effects drive coevolution in mutualistic networks. *Nature*, 550, 511–514.

Hastings, A. (2004). Transients: the key to long-term ecological understanding? *Trends in ecology & evolution*, 19, 39–45.

Holyoak, M., Leibold, M.A. & Holt, R.D. (2005). *Metacommunities: spatial dynamics and ecological communities*. University of Chicago Press.

Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography (MPB-32)*. vol. 32. Princeton University Press.

Knight, T.M., McCoy, M.W., Chase, J.M., McCoy, K.A. & Holt, R.D. (2005). Trophic cascades across ecosystems. *Nature*, 437, 880–883.

Krumbeck, Y., Yang, Q., Constable, G.W. & Rogers, T. (2021). Fluctuation spectra of large random dynamical systems reveal hidden structure in ecological networks. *Nature Communications*, 12, 1–14.

Levine, J.M., Bascompte, J., Adler, P.B. & Allesina, S. (2017). Beyond pairwise mechanisms of species coexistence in complex communities. *Nature*, 546, 56–64.

Levins, R. (1974). Discussion paper: The qualitative analysis of partially specified systems. *Annals of the New York Academy of Sciences*, 231, 123–138.

Levins, R. & Lewontin, R. (1982). *Dialectics and Reductionism in Ecology*, Springer Netherlands, Dordrecht, pp. 107–138.

Lewontin, R. (1996). *Biology as ideology: The doctrine of DNA*. House of Anansi.

Liautaud, K., Nes, E.H.v., Barbier, M., Scheffer, M. & Loreau, M. (2019). Superorganisms or loose collections of species? A unifying theory of community patterns along environmental gradients. *Ecology Letters*, 22, 1243–1252.

Loreau, M. (2010). From populations to ecosystems. In: *From Populations to Ecosystems*. Princeton University Press.

Loreau, M. (2020). The ecosystem. *Unsolved problems in ecology*, pp. 218–224.

Lovelock, J.E. & Margulis, L. (1974). Atmospheric homeostasis by and for the biosphere: the gaia hypothesis. *Tellus*, 26, 2–10.

Lyapunov, A.M. (1892). The general problem of the stability of motion. *Doctoral dissertation, University of Kharkov.*

May, R.M. (1972). Will a large complex system be stable? *Nature*, 238, 413–414.

May, R.M. (1973). *Stability and Complexity in Model Ecosystems.* Princeton University Press.

Menge, B.A. (1995). Indirect effects in marine rocky intertidal interaction webs: Patterns and importance. *Ecological Monographs*, 65, 21–74.

Montoya, J., Woodward, G., Emmerson, M.C. & Solé, R.V. (2009). Press perturbations and indirect effects in real food webs. *Ecology*, 90, 2426–2433.

Montoya, J.M., Pimm, S.L. & Sole, R.V. (2006). Ecological networks and their fragility. *Nature*, 442, 259–264.

Novak, M., Yeakel, J.D., Noble, A.E., Doak, D.F., Emmerson, M., Estes, J.A., Jacob, U., Tinker, M.T. & Wootton, J.T. (2016). Characterizing species interactions to understand press perturbations: what is the community matrix? *Annual Review of Ecology, Evolution, and Systematics*, 47, 409–432.

Orr, J.A., Piggott, J.J., Jackson, A.L. & Arnoldi, J.F. (2021). Scaling up uncertain predictions to higher levels of organisation tends to underestimate change. *Methods in Ecology and Evolution*, 12, 1521–1532.

Patten, B.C. (1982). Environs: Relativistic elementary particles for ecology. *The American Naturalist*, 119, 179–219.

Pimm, S.L. (1984). The complexity and stability of ecosystems. *Nature*, 307, 321–326.

Pires, M.M., O'Donnell, J.L., Burkle, L.A., Díaz-Castelazo, C., Hembry, D.H., Yeakel, J.D., Newman, E.A., Medeiros, L.P., de Aguiar, M.A. & Guimarães Jr, P.R. (2020). The indirect paths to cascading effects of extinctions in mutualistic networks.

Puccia, C.J. & Levins, R. (1991). Qualitative modeling in ecology: loop analysis, signed digraphs, and time averaging. In: *Qualitative simulation modeling and analysis.* Springer, pp. 119–143.

Puccia, C.J. & Levins, R. (2013). *Qualitative modeling of complex systems.* Harvard University Press.

Reed, M., Simon, B., Simon, B. & Simon, B. (1972). *Methods of modern mathematical physics.* vol. 1. Elsevier.

Roy, F., Biroli, G., Bunin, G. & Cammarota, C. (2019). Numerical implementation of dynamical mean field theory for disordered systems: Application to the lotka–volterra model of ecosystems. *Journal of Physics A: Mathematical and Theoretical*, 52, 484001.

Sanchez, A., Bajic, D., Diaz-Colunga, J., Skwara, A., Vila, J.C. & Kuehn, S. (2022). The community-function landscape of microbial consortia.

Schmitz, O.J. & Suttle, K.B. (2001). Effects of top predator species on direct and indirect interactions in a food web. *Ecology*, 82, 2072–2081.

Sharkey, K.J. (2017). A control analysis perspective on katz centrality. *Scientific reports*, 7, 1–8.

Simmons, B.I., Cirtwill, A.R., Baker, N.J., Wauchope, H.S., Dicks, L.V., Stouffer, D.B. & Sutherland, W.J. (2019). Motifs in bipartite ecological networks: uncovering indirect interactions. *Oikos*, 128, 154–170.

Soberón, J. (2007). Grinnellian and eltonian niches and geographic distributions of species. *Ecology letters*, 10, 1115–1123.

Tilman, D. (1994). Competition and biodiversity in spatially structured habitats. *Ecology*, 75, 2–16.

Trefethen, L.N. & Embree, M. (2020). *Spectra and pseudospectra*. Princeton university press.

Wootton, J.T. (1994). The nature and consequences of indirect effects in ecological communities. *Annual Review of Ecology and Systematics*, 25, 443–466.

Yodzis, P. (1988). The indeterminacy of ecological interactions as perceived through perturbation experiments. *Ecology*, 69, 508–515.

Yodzis, P. (2000). Diffuse effects in food webs. *Ecology*, 81, 261–266.

Supplementary Material

S1: On the arbitrary bounds of communities

By ‘community’, we mean a set of interacting species, which could represent part or all of an ecosystem (Levins, 1974). It is most likely that the community does not exhaust the whole ecosystem in which it is embedded (in fact the latter would be hard to define). An ecologist might consider a community of plant species in a grassland, knowing of course that this grassland also contains insect species, fungi, bacteria, birds and larger herbivores that do have an influence on the plants. Those unobserved species (or abiotic compartments) will mediate the interactions between the species that we do consider explicitly, so that the interaction between two species is not an intrinsic characteristics of these populations, but is context-dependent. Thus, each interaction term is a property of the system, in a given state, rather than a fixed property of a given species pair.

The notion of ”carrying capacity” –the abundance of a species if alone– must also be understood as context-dependent and not at all intrinsic to the focal species (Loreau, 2010). Carrying capacity changes as we change the scope of the community we are considering. If we only consider a single species, it must have a positive carrying capacity, otherwise it would be absent. But this carrying capacity might be an emergent feature of the myriad other species and abiotic compartment that we did not consider. As we increase the scope of the community under consideration, carrying capacity might become more intrinsic, but will most likely also decline. In the limit of all organisms on Earth, the notion itself ceases to make sense.

S2: Lotka-Volterra community assembly

Following is a description of the simulations and calculations used to make the figures in the main text. This includes the creation of model communities, simulations of their equilibrium and temporal dynamics, and calculation of different measures based on these simulations.

The presentation of signatures of collective integration in Figures 2-5 are all based on the same set of model communities. This is a set of generalized Lotka-Volterra communities with random carrying capacities and interactions. The equations simulated are given by eq. 21

$$\frac{dN_i}{dt} = N_i(K_i + \sum_j A_{ij}N_j) \tag{21}$$

We consider a gradient of interaction strength, with 50 different values of average interaction strength along the gradient, each with 100 random communities, making up 5000 communities in total. Each community starts with 50 species (stronger interaction leads some species to extinction in the equilibrium state), and we set 80% of the interactions to zero, so that we have a sparse interaction matrix. For defining the interaction matrix, we follow Barbier *et al.* (2018) and define the three parameters of random interactions as $\mu = -y; \sigma = y; \gamma = -1$, where y is the above-mentioned gradient, with values ranging between 0.02 and 1.0, in uniform intervals.

This translates to the following relations. The carrying capacities K_i are taken from a random uniform distribution centered on 1, with a standard deviation of 0.2 (within the span of $1-\sqrt{3}/5$ to $1+\sqrt{3}/5$). The interactions matrix A_{ij} is taken from a Gaussian distribution, except for the diagonal which is set to -1. For the off diagonal values, we constrain the values to keep three conditions: i) asymmetrical interactions, so that $A_{ij} = -A_{ji}$; ii) we use a sparsity of $\rho = 0.8$, i.e. 80% of of the interactions values are set to 0; iii) the average and standard deviation of the interaction values (including the values set to 0, but not the diagonal values) are $-y \cdot S$, $-y \cdot \sqrt{S}$, respectively ($S = 50$ is the number of species in the community).

These constraints are achieved by building a matrix in the following order: 1) take a matrix A_1 of size S from a random Gaussian distribution; 2) Get a new matrix by subtracting the original matrix from its transpose: $A_2 = A_1 - A_1^T$; 3) normalize the matrix by its standard deviation: $A_3 = A_2/std(A_2)$; 4) use the normalized matrix to get a certain average and standard deviation: $A_4 = A_3 \cdot y \cdot \sqrt{S(1-\rho)} - y \cdot S(1-\rho)$; 5) Choose at random a fraction of ρ of the upper triangualar part of the matrix, and set it to 0, as well as its lower triangualar counterpart.

This gives us asymmetrical interactions between species ($\gamma = -1$), and interactions that are increasingly negative (due to μ) and at the same time more varied (due to σ). Using this definitions, along the gradient we find communities with a collectivity parameter ϕ ranging between 0 and 2, allowing us to showcase different aspects of low and high collective integration in communities.

For each of the 5000 communities, we use the carrying capacity vector K_i and interaction matrix A_{ij} to assemble a community. We run a simulation until reaching an equilibrium (as detailed in the next subsection) with the starting conditions that all 50 species abundances equal 1. Depending on the location along the gradient, all species survive for low values of y, or at least 30 species survive at high y values. For any given community, we consider only the extant species for calculations and presentation (e.g. ϕ is calculated for the A_{ij} matrix where all rows and columns of extinct species are taken out).

S2.1 Simulations

The simulations of the ordinary differential equations (ODEs) given by eqs. 1 were preformed using the ode45 solver of Matlab (R2019a). To reach an equilibrium we start with a simulation time span of $T = T_0$, and iteratively run a simulation for longer time spans of T (doubling T) until the equilibrium condition defined below is reached, or until we reach a maximal simulation time T_{max}. The equilibrium condition is that at the last quarter of the simulation the maximal change in species abundance is below the threshold z, scaled by the time span T, $max(abs(N_i(T) - N_i(0.75T))) < 2zT$. We choose the numerical parameters $T_0 = 10^2, T_{max} = 10^4, z = 10^{-6}$, ensuring that the abundances no longer change in any noticeable way once reaching the threshold.

Presentation of Signatures

We describe here in detail the calculations and simulations used for the presentation of signatures of collective integration in Fig. 2-5.

Effective connectance

The signature of effective connectance is defined using aratio between two Shannon diversity index, applied on two matrices that describe the interactions within the community. We use the set of communities as described in the previous subsections, and make use of two matrices:the direct interactions matrix A_ij , and the net interactions matrix, $V_{ij} = (1-A_ij)^{-1}$. To calculate the effective connectance we use the ratio between the Shannon diversity of these matrices, calculated as follows.

For each matrix we calculate a diversity of interactions by: 1) taking all interactions except self-interactions (i.e. diagonal values), giving us $S(S-1)$ values; 2) computing the prevalence of each unique value as a fraction $\rho = x/S(S-1)$, where x is the number of occurrences of each value; 3) calculating the Shannon diversity on these values as $H = -\rho \cdot log(\rho)$. We then define effective connectance as the ratio of the two indices $\frac{H(A)}{H(V)}$. Note that for our definition of unique values, we round all values by 10^-3 (e.g. all values between 0 and 10^-3 are considered the same, all values between 10^-3 and $2*10^-3$ are considered the same, and so on), as otherwise all random values will be effectively unique and the definition of Shannon diversity would be rendered meaningless. The result of this calculation is used for the y-axis measure of Fig. 2a. For Fig. 2b-d we show the histograms of off-diagonal values for both A_ij and V_ij, for three specific communities along the y gradient.

Perturbation depth

The signature of perturbation depth is tied to the response of the community to an external perturbation, and we therefore calculate it using a dynamical simulation (as described in the previous subsections). For a given community at its equilibrium abundance N_j^*, we consider S different scenarios, in each of which we eliminate one species from the community, and observe the effect on the the remaining $S-1$ species. We simulate the community dynamics until it reaches a new equilibrium, and then measure the change in abundance for all $S-1$ species, $N_{j\setminus i}^*$. For this community, we also note the link distance $d(i,j)$ between two species i and j, where a distance of 1 is for directly interacting species, distance 2 is for species that do not directly interact but that both interact directly with the same third species, and so forth. We these ingredients we calculate the perturbation depth as:

$$\mathrm{PD} = \frac{1}{S}\sum_i \frac{\sum_j d(i,j)\left|N_{j\setminus i}^* - N_j^*\right|}{\sum_j \left|N_{j\setminus i}^* - N_j^*\right|} \tag{22}$$

Note that in these summations over j we go through all species except for the removed species, whereas the summation over i is essentially over different removal experiments. The result of this calculation is used for the y-axis measure of Fig. 3a. For Fig. 3b we calculate the relative impact of the perturbation of three groups of species, the ones directly interacting with the removed species (red), the one that interact with a directly interacting species (green), and the rest of the species in the community that are at a distance of 3 or more (blue). This relative impact is the ratio between the change in abundance within the group, and the change in abundance for all the species in the community (except the removed one), and all of this is averaged over the different removal experiments.

$$\mathrm{RI}_k = \frac{1}{S} \sum_i \frac{\sum_j \delta_{d(i,j),k} \left| N^*_{j\setminus i} - N^*_j \right|}{\sum_j \left| N^*_{j\setminus i} - N^*_j \right|} \tag{23}$$

Temporal unpredictability

The signature of temporal unpredictability relates to the difference between short-term and long term response, and as such we use dynamical simulations both up to equilibrium, and also within a short timescale.

We look at how the community responds to perturbations by changing the carrying capacity of each species at random: $\tilde{K}_i = K_i + \zeta$ where ζ is chosen at random from a uniform distribution centered around 0, ranging from -0.25 to 0.25. We then run a simulation twice, first over a short time span $\Delta T = 0.5$, and then until the system reaches a new equilibrium $\tilde{N}_i$. From the first simulation, for each species separately, we extract 10 time points (regularly spaced) within ΔT, and fit the points to a linear function. We use the slope α_i from the fit to extrapolate what the long term response is expected to be: $R_S = \alpha_i/(K_i N_i)$. This is compared to the actual long term response $R_L = \tilde{N}_i - N_i$, giving us our signature of temporal unpredictability, as defined by eq. 24.

$$\mathrm{TU} = 1 - \frac{cov(R_S, R_L)}{std(R_S)std(R_L)} \tag{24}$$

The result of this calculation is used for the y-axis measure of Fig. 4a. For Fig. 4b-c we also demonstrate the dynamics of three communities along the y gradient. Fig. 4b-c we show the actual long tern dynamics (solid lines), compared with the predicted long term dynamics, extrapolated from the short term (dotted line). The extrapolated curves shown are: $\Delta\tilde{N}_i = (1 - exp(N_i K_i t)) \cdot N_i K_i/\alpha_i$.

Biotic contribution to the realized niche

Biotic contribution (BC) measures how much the abundances are determined by the community as a whole, rather than by the abiotic conditions alone. We measure it using the relative yield of species $\eta_i = N_i/K_i$, where we use the equilibrium abundances N_i. This gives us the signature, used for the y-axis of Fig. 5a, as

$$\mathrm{BC} = \frac{\sum_i (\eta_i - 1)^2}{\sum_i \eta_i^2} \tag{25}$$

For Fig. 5b-d, we use for the x-axis the carrying capacities K_i (blue), and the term $< K > \sum_j V_i j$ which represents the effects of the community on the species (red), where we sum over the matrix of net interactions $V = (1 - A)^{-1}$.

S3: Spectral radius and Frobenius norm of real zero-trace matrices

Lemma 1. *Let $A \in M_n(\mathbb{R})$ (a real square matrix) such that $TrA = 0$. Let $\lambda_0, \lambda_1, ...\lambda_{n-1} \in \mathbb{C}$, $i = 0, ..., n-1$ are its n eigenvalues (possibly repeated), ordered in decreasing order of*

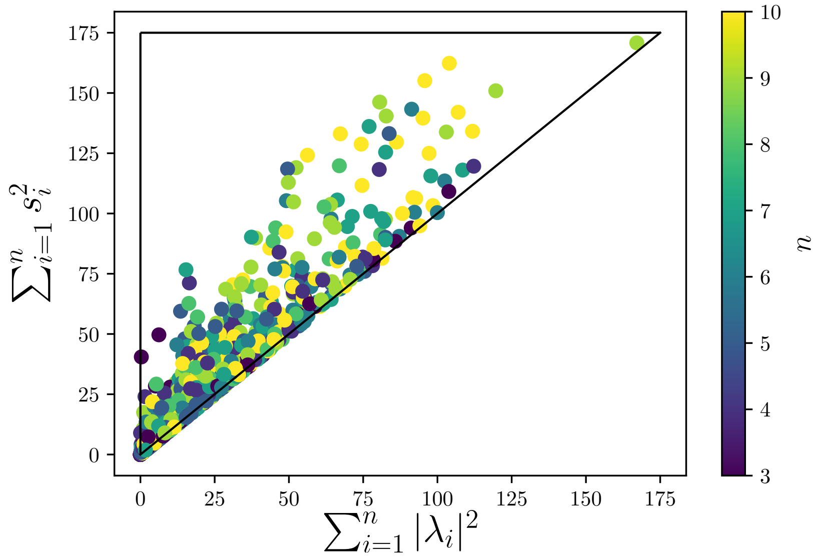


Figure S1.1: **Test of conjecture 4:** for a wide array of randomly generated matrices of size $n \times n$ with $n = 3, ..., 10$. We could never find any matrix for which the inequality $\sum_{i=1}^{n} |\lambda_i|^2 \leq \sum_{i=1}^{n} s_i^2$ (and we know it holds for $n = 2$).

modulus, so $|\lambda_0| \geq |\lambda_1| \geq ... \geq |\lambda_{n-1}|$*. It then holds that*

$$|\lambda_0|^2 \leq \left(1 - \frac{1}{n}\right) \sum_{i=0}^{n-1} |\lambda_i|^2$$

Proof. If λ_0 has an imaginary part, then it comes as a conjugate pair with $\lambda_1 = \bar{\lambda}_0$, and the two have equal modulus. In this case $\sum_{i=0}^{n-1} |\lambda_i|^2 \geq 2|\lambda_0|^2$. Note that that $\frac{1}{2} \leq 1 - \frac{1}{n}$ for all $n \geq 2$, so in this case the upper bound holds. Now suppose λ_0 is real. Due to the fact that $\mathrm{Tr}A = 0$, and that A is real, we have $\sum_{i=0}^{n-1} \Re\lambda_i = 0$. The maximal configuration for the rest of eigenvalues is that they all have the same real part, equal to $-\frac{\lambda_0}{n-1}$. In this case $\sum_{i=0}^{n-1} |\Re\lambda_i|^2 = \frac{n}{n-1}|\lambda_0|^2$. This shows that, when λ_0 is real, it always hold that $|\lambda_0|^2 \leq (1 - \frac{1}{n}) \sum_{i=0}^{n-1} |\Re\lambda_i|^2$ which is obviously smaller that $(1 - \frac{1}{n}) \sum_{i=0}^{n-1} |\lambda_i|^2$. □

Definition 2. Let $\tilde{M}_n(\mathbb{R})$ be the non-empty subset of $M_n(\mathbb{R})$ comprised of matrices A such that $\sum_{i=0}^{n-1} |\lambda_i|^2 \leq \sum_{i=0}^{n-1} s_i^2$, where $s_0^2 \geq s_1^2 \geq ... \geq s_{n-1}^2 \geq 0$ are the eigenvalues of the positive-definite matrix $A^\top A$.

For normal matrices, $|\lambda_i| = s_i$, so clearly $\tilde{M}_n(\mathbb{R})$ is not empty as it contains normal matrices.

Lemma 3. *For any* $A \in \tilde{M}_n(\mathbb{R})$ *such that* $TrA = 0$*, it holds that*

$$|\lambda_0| \leq \sqrt{1 - \frac{1}{n}} ||A||_F$$

where $||A||_F = \sqrt{\sum_{i,j} A_{ij}^2}$ *is the Frobenius norm of* A*.*

Proof. From lemma 1 we have that $|\lambda_0|^2 \leq \left(1 - \frac{1}{n}\right) \sum_{i=0}^{n-1} |\lambda_i|^2$ and if the matrix is in $\tilde{M}_n\left(\mathbb{R}\right)$ then it follows that $|\lambda_0|^2 \leq \left(1 - \frac{1}{n}\right) \sum_{i=0}^{n-1} s_i^2$. On the other hand, we may note that $\sum s_i^2 = \mathrm{Tr} A^\top A = \sum_i (A^\top A)_{ii} = \sum_i \sum_j A_{ij}^2 = ||A||_F^2$. We get the equivalent expression but written in terms of May's complexity measure by noting that $\mathcal{C} = ||A||_F^2/\sqrt{n}$. □

Conjecture 4. *$\tilde{M}_n\left(\mathbb{R}\right)$ is a generic set. This means that, if there exists a real square matrix not in $\tilde{M}_n\left(\mathbb{R}\right)$, arbitrarily small perturbations of this matrix will be.*

We can show by a direct computation that $\tilde{M}_2\left(\mathbb{R}\right) \equiv M_2\left(\mathbb{R}\right)$. In higher dimensions extensive simulations (some shown in figure) support the idea that it is, if not impossible, very unlikely that a real matrix does not belong to $\tilde{M}_n\left(\mathbb{R}\right)$.