bioRxiv preprint doi: https://doi.org/10.1101/145862; this version posted June 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Generic assembly patterns in large ecological communities

Matthieu Barbier[1,*], Jean-François Arnoldi[1], Guy Bunin[2] and Michel Loreau[1]

June 4, 2017

Abstract

Ecological communities have mainly been investigated theoretically in two ways: piecewise, a few species at a time; or as complex networks, simulated in exhaustive detail. But our empirical knowledge of networks is limited, and the space of simulation models and parameters is mindbogglingly vast. We show that a large fraction of that space of possibilities exhibits generic dynamics, which can be predicted from a single minimal model. To demonstrate this, we consider a wide array of ecological models, from resource competition to predation and mutualism, known to display very different behaviors for a few species. We simulate large communities, and show that equilibrium diversity, functioning and stability can often be predicted analytically from only four broad statistical properties of the community. Our approach provides a convenient framework for exploring generic patterns in ecosystem assembly and quantifying the added value of detailed models and measurements.

1 Introduction

As famously pictured by Darwin's "entangled bank", ecological communities form large and intricate networks of dynamical interdependencies between their constituent species and abiotic factors. The richness and variety of natural systems has long been a source of wonder, but understanding the mechanisms that shape their diversity, stability and functioning is also a pressing challenge.

Many of the intuitions we employ to "disentangle the bank" have come from the study of simplified patterns of interaction, involving a few species with distinctive traits (e.g. a predator and a prey, two parasites and their host). There are two main ways that these results have been used to understand larger communities. First, such simple structures can remain *globally* predictive: each species in a simplified model can stand for an entire group, for instance generalizing from predator-prey dynamics to whole trophic levels. Second, they may remain *locally* predictive: a more complex network, while not so neatly organized, might still be decomposed into different patterns (network motifs [1, 2] or feedback loops [3]) and understood piece by piece.

However, the way that many small-scale motifs come together to drive the dynamics of large communities is often unclear. The next step for ecological theory was to investigate complex networks of interacting species. This often relied on extensive numerical simulations, randomly drawing species traits from empirically-inspired distributions (e.g. [4]). Current computational power allows exploring levels of complexity that were unapproachable only a few decades ago, and ecologists have taken advantage of this fact to address numerous questions about the properties of communities assembled from a large pool of interacting species. But the parameter space in these simulations is fatally vast, with many possible choices for species traits, interactions, and even dynamical equations. Thus, although computers can now simulate realistic degrees of complexity, the absence of

[1]Centre for Biodiversity Theory and Modelling, Theoretical and Experimental Ecology Station, CNRS and Paul Sabatier University, 09200 Moulis, France.

[2]Department of Physics, Technion-Israel Institute of Technology, Haifa 3200003, Israel.

[*]To whom correspondence should be addressed. E-mail: contact@mrcbarbier.org

bioRxiv preprint doi: https://doi.org/10.1101/145862; this version posted June 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

empirically well-established structures and parameter values means that there is a discouragingly vast ocean of possibilities to explore. For each simulation, there could be a contradictory prediction given different yet plausible parameters and assumptions.

Against this vertiginous perspective, we show here that large communities can in fact exhibit simple and generic behavior, agnostic to most of these details. The crucial intuition is that, when combining many small structures, the outcome will fall between two limits. As we noted above, there can be systems where a simple, coherent global structure emerges. But in many cases, each species will be involved in various patterns that drive its dynamics in different directions, and cancel each other out. The latter case is what we call a *disordered* interaction network [5]. Its constituent motifs lose their unique signature; the more numerous they are, the more they interfere with each other. Instead, ecosystem properties become sensitive only to a few aggregated features of the species pool, leading to a drastic reduction of the space of possibles.

Our purpose in this article is to demonstrate this property of simplicity arising from disorder, and discuss when it is relevant to ecological community assembly. We borrow an array of models from the literature, with different dynamics and structures, covering the main types of ecological interactions: mutualism, competition and predation. For a few strongly-coupled species, these models are known to lead to wildly different outcomes, from competitive exclusion to predator-prey cycles. However we show that, when interactions are distributed over more species, the outputs of all these models can become predictable from a much simpler, more generic "reference model". This random Lotka-Volterra model covers a class of possible dynamics, rich enough to encompass the outcomes of the other models, yet simple enough to permit the exhaustive analysis of essential community properties [6] including species richness, abundance distribution and stability.

The reference model is parametrized by only four simple statistics computed from species interactions, growth rates and self-regulation. Armed with this minimal information, it reproduces the results of various complex simulations, suggesting that the main mechanisms controlling their dynamics are not model-specific. This is a powerful simplification, which allows us to identify the role of various modelling choices, make sense of contradictory predictions obtained in different simulations, or simply use the reference model as a starting point in future investigations to tease out other generic patterns in assembly. The idea that few parameters can drive properties of complex systems is not new in ecology: May famously related the stability of large randomly communities to only a few statistical properties of the interactions [7, 8]. Because we explicitly consider the assembly process, our approach goes further: first, it accounts for the fact that, even starting from a random pool, population dynamics produce non-random patterns in assembled communities that ensure their stability [9]; second, our results extend to other ecosystem properties, such as functioning and diversity.

The reference model cannot always predict the equilibria of other models, but such discrepancies also produce novel insights. Far from being haphazard, they occur when the community possesses a coherent global structure, such as a trophic or competitive hierarchy. In fact, the two limits of randomness and structure are not exclusive, but complementary: a large part of what makes complex ecosystems different from simplified models resides in the disordered component, but it may have to be combined with essential information about community structure.

Our results are structured as follows: first, we demonstrate that the equilibria of complex ecological dynamics can be predicted by a simple reference model, parametrized by a few statistical properties of the species pool. Second, we argue that these generic parameters provide a powerful tool to reduce the parameter space of other models, and efficiently explore and compare their predictions. We show that certain, but not all, network structures lead to less generic phenomena that diverge from reference model predictions. Finally, we demonstrate how the latter can be extended to capture these structures in the simplest possible way, and propose this combination of disordered dynamics and simple structure as a way forward in the task of understanding complex ecological communities.

bioRxiv preprint doi: https://doi.org/10.1101/145862; this version posted June 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

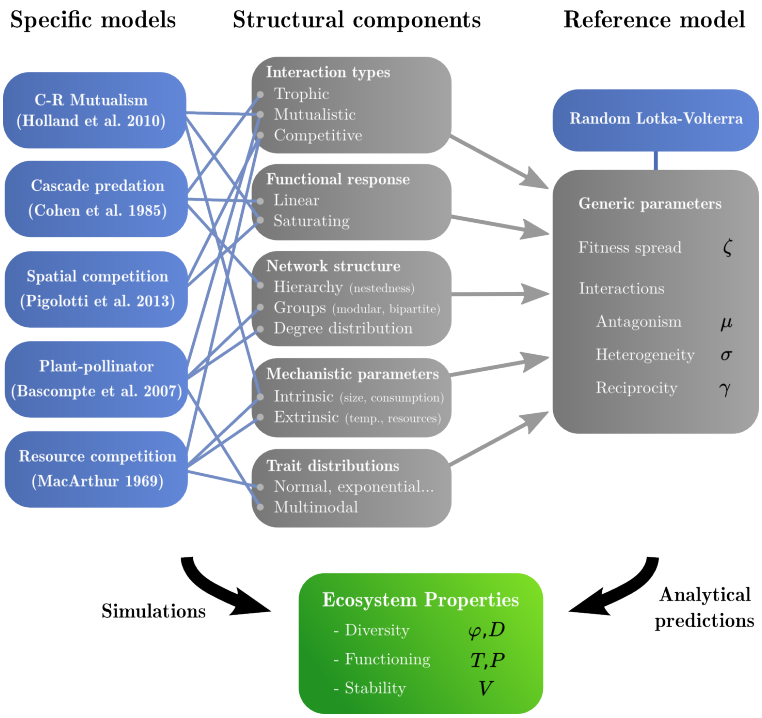

Figure 1: Comparison scheme used throughout this article. On the left, various specific ecological models from the literature, some of which are used as examples below. They correspond to various choices of model ingredients, among the main categories detailed in Box 1 (not all connections are represented here). Numerical simulations of these models (and other combinations of ingredients) allow us to find their properties at equilibrium. On the right, the reference Lotka-Volterra model with its four control parameters, whose values can be obtained by "translation" from the ingredients of other models. This reference model is easily simulated, and it is also solved analytically, so that its equilibrium features are known and can be compared to simulations.

bioRxiv preprint doi: https://doi.org/10.1101/145862; this version posted June 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Box 1: Reference model and community properties

Dynamics The reference random model follows the widely-used Lotka-Volterra dynamics:

$$\frac{d}{dt}N_i = \frac{r_i}{K_i}N_i\left(K_i - N_i - \sum_{j\neq i}^{S}\alpha_{ij}N_j\right) \quad (1)$$

where S is the number of species in the pool, K_i the carrying capacity of species i (if $K_i > 0$, else density-dependent mortality), $\alpha_{ij}N_j$ the interaction with species j, and r_i the intrinsic growth rate. Solving for equilibrium $dN_i/dt = 0$,

$$N_i = K_i - \sum_{j\neq i}^{S}\alpha_{ij}N_j. \quad (2)$$

where we see that the growth rate r_i plays no explicit role. Only the species for which (2) allows $N_i > 0$ can survive together in an assembled state. We consider the equilibrium obtained after all possible invasions took place[a].

Parameters As recently shown [6], if α_{ij} and K_i are drawn at random, the equilibrium properties are controlled by four statistics of the pool of invaders:

$$\zeta^2 = \left\langle K_i^2\right\rangle - \left\langle K_i\right\rangle^2, \qquad \mu = S\left\langle\alpha_{ij}\right\rangle,$$
$$\sigma^2 = S(\left\langle\alpha_{ij}^2\right\rangle - \left\langle\alpha_{ij}\right\rangle^2), \qquad \gamma = \frac{S}{\sigma^2}\left(\left\langle\alpha_{ij}\alpha_{ji}\right\rangle - \left\langle\alpha_{ij}\right\rangle^2\right), \quad (3)$$

Features Ecosystem functioning was characterized by the total biomass T and total productivity from external resources P, defined from species abundances N_i (in units of biomass) and growth rates r_i. Diversity was represented by two quantities: ϕ the fraction of species in the invader pool that survive in the assembled state, and D the inverse of the Simpson index [10]. Among the many dimensions of ecological stability [11], we focused on one measure of dynamical stability frequently employed in empirical studies: variability V, the variance in time of species abundance due to stochastic perturbations [12, 13, 14] (see details in Supplementary Materials). In summary:

$$T = \sum_i N_i, \quad P = \sum_i r_iN_i, \quad \phi = \sum_i \Theta(N_i), \quad (4)$$

$$D^{-1} = \sum_i\left(\frac{N_i}{T}\right)^2, \quad V = \sum_i\frac{\mathrm{Var}(N_i(t))}{S}. \quad (5)$$

with $\Theta(x) = 1$ if $x > 0$, 0 otherwise. Finally, all these quantities were combined in a single metric of relative error between simulations and reference:

$$\mathrm{Error} = \frac{1}{5}\sum_{x\in\{T,P,\phi,D,V\}}\left|\frac{x_{\mathrm{simulated}}}{x_{\mathrm{reference}}} - 1\right| \quad (6)$$

[a]It is possible to show that there is a critical degree of heterogeneity σ_c under which there is a unique, globally stable equilibrium reached by any invasion sequence, and above which there is multistability [6]. The analytical results are exact in the single-equilibrium regime, but still approximately valid in the multistable region.

2 Methods

2.1 Overview

We compared the equilibria of complex simulated communities to analytical predictions from a simple reference Lotka-Volterra model, following a procedure illustrated in Fig. 1. This was done in two steps:

(1) We selected models from the literature, and for each, generated many pools of species. We then simulated their dynamics, allowing invasions from any species of the pool at any time, until a stable, uninvadable equilibrium was reached. We measured important properties of this equilibrium: its diversity, functioning and dynamical stability.

(2) We translated each model into the generic parameter space of the reference Lotka-Volterra model. For each of the species pools generated earlier, we computed four statistical metrics – μ, σ, γ and ζ – which characterize the interactions and growth rates of potential invaders. We then inserted these four parameters into the analytical predictions for equilibrium features of the reference model [6, 15], and finally compared these predictions with the simulation results of the initial, more complex models.

2.2 Simulation models

We selected from the literature a wide array of models of community dynamics [16, 17, 18, 19, 20, 4, 21, 22], with starkly different ecological implications. Their common trait was having been studied for their equilibrium properties – we ignored stochastic models that would not give rise to stable assembled communities [23, 24, 25]. These models varied in an immense space of possibilities, from which we extracted five main classes of structural components (represented in Fig. 1):

Interaction nature and functional response We allowed for competitive, mutualistic, trophic and mixed interactions, and limited ourselves to linear or saturating functional response. This saturation can come from many factors: limited provision or diminishing returns for mutualistic services [19], finite

bioRxiv preprint doi: https://doi.org/10.1101/145862; this version posted June 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

handling time for predators [26] or competition with close neighbors in space [27].

Network structure We constrained the interactions to exhibit various complex network properties, which have been well studied in ecology and beyond [28]. One such property was nestedness, as exemplified by the cascade model for predation [18], or the competition-colonization model [17]: the first species consumes (or outcompetes) all the others, the second species consumes (or outcompetes) all but the first one, and so on. Another property was graph partitioning: many mutualistic networks in particular are bipartite, e.g. plant-pollinator networks [29]. We could also generate the interaction network from a classic ensemble: Erdos-Renyi random graphs [30] and Barabasi-Albert scale-free graphs [31]. These ensembles lead to very different network properties, such as degree distributions, which are often argued to play a fundamental role in network structure [32].

Parametrization and trait distributions In many models, interactions and carrying capacities were not drawn directly but derived mechanistically from underlying species traits and environmental factors, such as body size or temperature [4, 33]. Since there is no exhaustive way to explore these system-specific choices, we focused on the important example of resource competition [16, 34] below. In addition, all these traits could be drawn from a variety of distributions, including: normal, exponential, beta and multimodal.

From this list of structural components, we performed the following simulations (details in SI):

Sweep through interaction structures We simulated 180 combinations of three main structural components: interaction nature, functional response, and network structure (starting from a random graph and progressively ordering it to achieve desired nestedness, partition or degree distribution) to evidence how they varied in their impact on community assembly. Some of these combinations were similar to models from the literature, including those shown in Fig. 1: spatial competition [27], consumer-resource mutualism [20] and cascade predation [18].

Mechanistic example: Resource competition We studied a model where structure arose from mechanistic assumptions: species competition through consumption of various resources [16, 34], here chosen discrete (e.g. prey species, spatial patches). If ξ_{ix} is the rate at which species i consumes resource x (among R), the effective competition between species i and j was given by $\sum_{x=1}^{R} \xi_{ix}\xi_{jx}$. Likewise, species carrying capacities and growth rates were parametrized by consumption rates, and by the availability of each resource.

Modular example: Plant-pollinator community We wished to showcase a complex community with multiple functional groups and mixed interactions, inspired by theoretical and empirical considerations [29, 35, 22]. Our example was comprised of two functional groups, plants and pollinators. Interactions within each group were competitive, while those between groups were mutualistic, with pollinators being obligate mutualists (no intrinsic growth). To test the importance of group structure, we then increased the probability of rewiring each of these interactions, ignoring group boundaries while respecting nestedness and other desired network properties.

Model comparison

Generic parameter space The reference random model is parametrized by simple statistical moments of the carrying capacities K_i and interactions α_{ij}. The variance in species' intrinsic fitness is captured by ζ. For any species, the *total* effect of its interactions with all its partners is distributed with mean μ and standard deviation σ. This makes μ a measure of how antagonistic the community as a whole is to one species, and σ a measure of how much this varies between species. The reciprocity parameter γ takes values in the range $[-1, 1]$: $\gamma = 1$ means that two partners affect each other identically, while $\gamma = -1$ means that reciprocal effects are maximally different. The fact that these simple statistics – mean, variance

bioRxiv preprint doi: https://doi.org/10.1101/145862; this version posted June 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and symmetry – are sufficient to analytically predict many simulation results below shows that disorder can erase the dynamical influence of any higher-order correlation structure.

Translation For simulation models with a linear functional response, the translation process was immediate since they already had Lotka-Volterra dynamics and only needed to be rewritten in the form of equation (1). Then, we computed four statistics of these matrices as defined in (3) to parametrize the reference model, effectively ignoring all additional structure.

In the case of models with a saturating functional response, interactions were similar to Lotka-Volterra at low abundances but plateaued at higher abundances. Since the reference model allowed us to compute the abundance distribution, and hence the likelihood of saturation, we could derive the effective distribution of interaction strengths in the system, and compute the same statistics (technical calculations in SI).

Finally, we used various community properties to compare the equilibria attained in simulations to predictions from the equivalent random model, which are entirely known analytically [6]. Whenever the simulation results and reference predictions concurred, we deduced that any additional information contained in the simulation models failed to change these outcomes.

3 Results

Following the scheme outlined in Fig. 1, we compared the equilibrium properties of the simulated models to analytical predictions from the random Lotka-Volterra reference model. In Fig. 2, we showcase some of the community properties listed in Fig. 1 in a particular example: a resource competition model (see Methods). We show the effects of varying the number of resources, R, and a measure of species heterogeneity: the variance of consumption rates, σ_ξ^2. Increasing R led to more diversity without affecting other properties, while increasing σ_ξ^2 improved stability and functioning. The reference random Lotka-Volterra

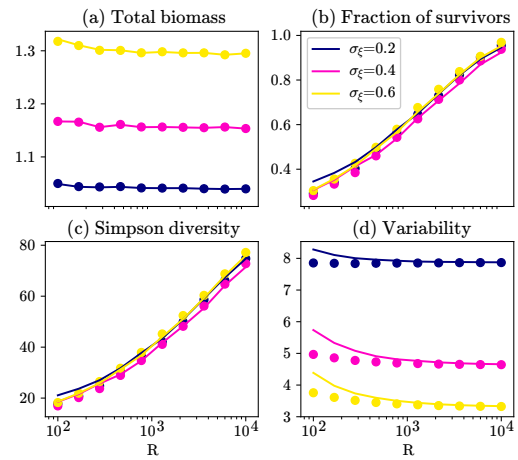

Figure 2: Various community properties predicted by the reference model: (a) Total biomass T, (b) Fraction of surviving species ϕ, (c) Simpson diversity D, (d) Temporal variability V. Simulation results (dots) were obtained from the resource competition model, varying the number of resources R and the heterogeneity of consumption rates σ_ξ (details in SI). They were then compared to theoretical predictions from the reference model (solid lines) parametrized only by the four generic properties of the species pool, see Fig. 1.

bioRxiv preprint doi: https://doi.org/10.1101/145862; this version posted June 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

model, lacking the underlying mechanistic structure, was simply parameterized by the four statistics μ, σ, γ and ζ defined in Box 1 in (3). It could then reproduce all these patterns *quantitatively*. In other words, R, σ_ξ^2 and other properties were important only inasmuch as they affected these four generic parameters.

We illustrate in Fig. 3 how different models visit this generic parameter space as we vary their own properties, focusing on the (μ, σ) plane. This projection in a common space allows to understand why and when various models can generate similar or different predictions. For instance, models can overlap: we constructed an example where the competition and predator-prey models correspond to the same values of all generic parameters, giving rise to communities with identical properties (in particular, species abundance distributions). These very different models thus indeed live in the same space of generic dynamics. However, they mostly occupy different regions of it, and this overlap is rare for all four parameters: for instance, resource competition tends to lead to higher μ but lower σ than predation, entailing lower total biomass but more stability.

Fig. 4 explores the role of network structures for competitive, predator-prey and mutualistic communities. We plotted the relative error of the analytical predictions compared with simulations as we added more and more structure to the interactions. Strikingly, mutualistic communities emerged as uniquely disordered. While network structure (nestedness, bipartition) has been extensively studied [28, 36], we found that it did not contribute to the large-scale community properties studied here, except indirectly by changing μ, σ and γ. This is reminiscent of previous results on "universality" in mutualistic communities [37]. The same cannot be said of predator-prey communities, where nestedness appears to play a large role, as it creates a trophic hierarchy of species. Counter-intuitively, competitive communities seem to be those where structure matters the most, even though they are the least studied in that respect. Nestedness (competitive hierarchy) has long been proposed as an important structural factor [38], but we found here that the largest deviation from reference occurred in the bipartite case: we even observed that one entire group could go extinct while

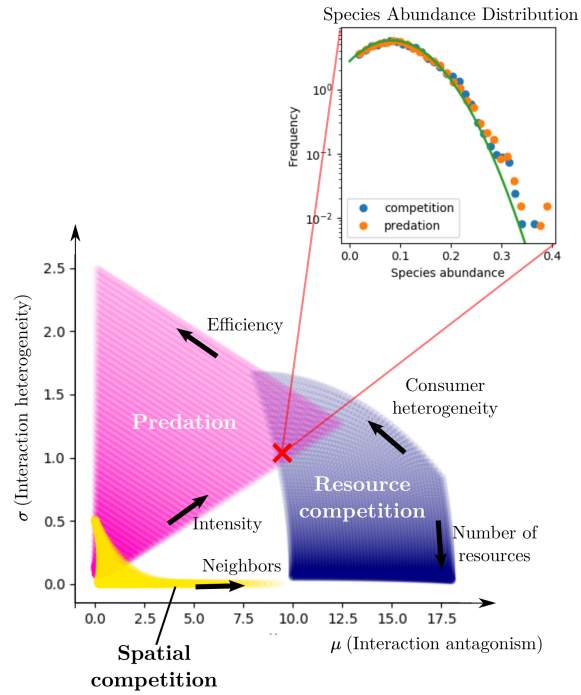

Figure 3: Equivalences and differences between communities. To explain how different models typically make distinct predictions, but can nevertheless overlap, we show where they lay once translated into the parameter space of the reference random model, following the comparison scheme outlined in Fig. 1. We illustrate in the (μ, σ) plane the effective parameter values corresponding to: the spatial competition model (yellow) with neighbor threshold $N_c \in [1, 100]$ and mean interaction $\tilde{A} \in [1, 20]$; predation (magenta) with intensity $m \in [0.1, 25]$ and biomass conversion efficiency $\epsilon \in [0, 1]$; and resource competition (blue) with number of resources $R \in [10^2, 10^4]$ and consumer heterogeneity $\sigma_\xi \in [0.1, 0.6]$. See SI for details on model-specific parameters. *Inset:* An example where a competitive community and a predator-prey community display identical species abundance distributions, corresponding to the parameter values marked by the red cross.

bioRxiv preprint doi: https://doi.org/10.1101/145862; this version posted June 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

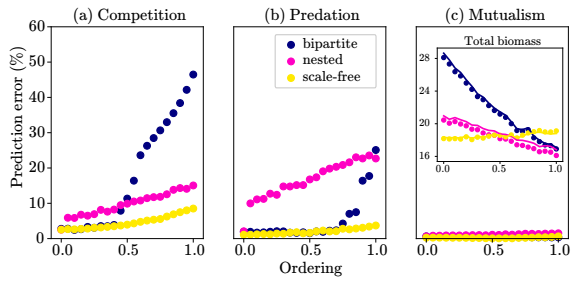

Figure 4: The role of network structure. (a)-(c) For each of the three main interaction types, the relative error (y-axis, between 0 and 60%) of the reference model against simulations, as a function of the degree of structure in the community (x-axis). Each set of symbols indicates a different network structural property: nestedness, partitioning (from complete to bipartite graph) and scale-free structure. The ordering parameter is $1-p$ with p the probability of rewiring any of these interactions at random; the fully disordered limit is always a random graph. We see that mutualistic communities are especially predictable from a disordered model, with less than 2% error regardless of their structure. *Inset:* Total biomass in mutualistic communities. We can see that predictability is not due to lack of change: ordering *does* seem to have strong effects on simulation results (symbols), but these effects are entirely predicted by the reference model (solid lines), i.e. they simply translate to changes in μ, σ and γ.

the other survived, suggesting competitive exclusion at the group level [39].

Following this indication of the importance of functional groups, the reference model was extended to allow for multiple groups, measuring interaction parameters μ, σ and γ within each group and between groups (see Methods). We illustrate in Fig. 5 the case of a plant-pollinator community with intra-group competition and inter-group mutualism. Our extended theory made good predictions for simulations with a functional group structure, even in the complex situation of intermediate levels of order, where the boundaries between groups were blurred. We discuss other extensions in SI, in particular to account for nestedness. This prospective result suggests the value of incorporating simple structure and disorder simultaneously in a single model, to tackle more complex communities while staying in a relatively small parameter space.

4 Discussion

Generic patterns can emerge in the assembly of large, heterogeneous ecological communities, that depend only on global statistical properties of the species pool, rather than on detailed model assumptions. Understanding these patterns provides powerful tools to efficiently explore complex models and assess the importance of structure. To show this, we compared assembled communities at equilibrium, using a variety of properties accounting for diversity, stability and functioning. We used the same reference model as a comparison point for all others, and suggested how that model could be extended to account for community structure.

Theoretical consequences

Our results have at least two consequences for the theory of complex ecological communities. Intricately detailed ecological models may make the same predictions as a much simpler model, entirely parametrized by the easily computed summary statistics described in (3). On one hand, this means that theoretical investigations aiming for generic patterns can

bioRxiv preprint doi: https://doi.org/10.1101/145862; this version posted June 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

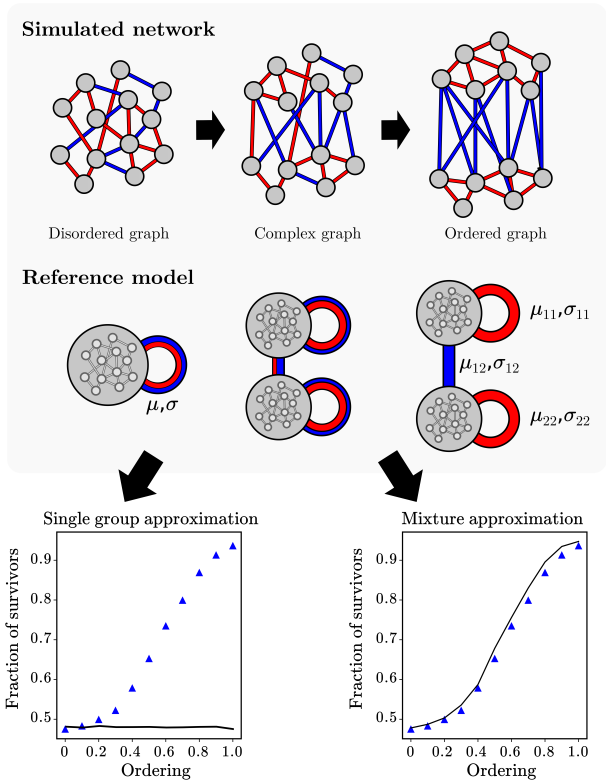

Figure 5: Mixture model: two functional groups with competitive and mutualistic interactions. (a) Cartoon of the model. In the ordered limit, all intra-group interactions were competitive and all inter-group interactions were mutualistic. The ordering parameter is defined as in Fig. 4. Mutualism is facultative for one group and obligate for the other. (b) Fraction of surviving species in the assembled community for the simulation model (symbols) against analytical predictons in the basic reference model (solid line) which cannot account for ordering. (c) Same data, but the reference model is modified to account for the existence of two groups. It successfully predicts community properties for any degree of ordering, even the complex intermediate case.

start from the reference model as a flexible and simple platform for exploring various ecological patterns. In addition, this generic character is not limited to a "mean-field" limit where all differences between species are erased statistically, as in previous statements on universality in ecological dynamics [40, 41, 37].

On the other hand, theorists interested in a more specific and complex model may benefit from comparing its predictions to that of the reference model. This is immediate as there is no fitting involved: the parameters of the reference model are simple statistics of the species pool, and can be inserted in the reference model's analytical solution to readily obtain predictions. Checking when these results align is a way of testing the "added value" of other structures and mechanisms. If outcomes differ, this suggests that some aspect of the simulation model's structure has large-scale dynamical consequences. If they are identical, then we can better understand the role of the model's parameters, by seeing how they translate into the reference parameters whose effect on equilibrium properties are easily understood. In particular, this can be useful to disentangle generic community dynamics from mechanistic parametrization, leading to a more modular approach: instead of putting all components together in a large simulation, we can see how certain mechanisms (e.g. temperature or size dependence in growth and interactions) affect the generic parameters μ, σ, γ and ζ, and therefore how they will most generically shape the assembly process.

The role and interpretation of the generic parameters

Whenever the reference Lotka-Volterra model manages to capture the properties of the assembled state produced by another model, we conclude that all the relevant properties of the community were correctly encapsulated in the four parameters of the reference model: the summary statistics of the initial species pool μ, σ, γ and ζ. Large μ means more antagonistic interactions, hence lower total biomass. If species differ more in fitness, due to intrinsic factors (ζ) or interactions (σ), we find lower coexistence and Simp-

bioRxiv preprint doi: https://doi.org/10.1101/145862; this version posted June 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

son diversity. Finally, more negative γ means greater coexistence and stability (see also Fig. S1 in SI).

These four parameters are very generic attributes, that are not tied to any specific interaction nature or community structure. Instead, they reflect a top-down perspective on the puzzle of assembling a stable community from a large pool of species. ζ indicates how species differ in their intrinsic ability to remain in the assembled state. μ indicates how much biomass is lost (or gained, if negative) on average due to interactions. σ represents how different the "pieces" of the puzzle are in their connections; as σ decreases, the pieces become more similar, hence easier to assemble and less likely to lead to diverse solutions. γ in turn represents how symmetrical these connections are: negative γ can be associated with trophic interactions, but also with strongly asymmetrical competition or mutualism. For example, we often found $\gamma < 0$ in the resource competition model. This asymmetry causes a negative feedback of a species on itself via its immediate interactions, which makes it easier to assemble the community, leading to stabler, more species-rich equilibria.

Empirical consequences

From an empirical perspective, the details of species interactions are often difficult to observe and measure. The fact that relevant ecological patterns may only depend on aggregated properties of the interactions is a positive message: we could make predictions that are robust to a lack of detailed information, and that rely on a minimal number of fitted or inferred parameters.

On the other hand, even when these aggregate, "disordered" properties do not suffice in themselves, we suggest that a complex community may be understood by a combination of simple structure and disorder. This means that empirical work focusing on a few exemplary species can effectively capture the main dimensions of order in an ecosystem, while most of the surrounding diversity contributes a diffuse, variable background that can be probed with its own metrics. Coherent global structure has indeed been found in large-scale empirical studies involving thousands of interactions [22]. This suggests that it is possible (as in the mixture model, Fig. 5) to combine bottom-up and top-down, local and global, species and ecosystem perspectives, without erasing the importance of heterogeneity in a community.

Implications for future work

We have shown that, even when we try to build a detailed picture of an ecological community, its collective dynamics can often be understood from a few large-scale properties, that do not strictly follow the intuitive categories of ecological mechanisms (such as interaction types or network structures). Our work offers an outlook on what complexity means in an ecological setting. Predicting the fate of a certain species at a given locale, e.g. for conservation, may require knowing about every important feedback within its environment, biotic and abiotic. But one rarely needs all these details at once to understand the aggregate properties of an ecosystem, or the fate of most species most of the time. Instead, the exhaustive study of ecological networks could also pave the way toward finding new dimensions of simplicity at the collective level.

The idea of "disorder" that we use here should not restrict our approach to models with random interactions. Purely random systems are in some sense special – the total absence of order is a peculiar feature in the infinite space of possible communities. Instead, the results shown here suggest that predictions from a random model do not strictly require randomness to hold: they only require that non-random motifs interfere and lose their influence due to the number and heterogeneity of interacting components. In such cases, the equilibrium is only shaped by generic assembly mechanisms that are also present in random systems. This has allowed us to use a single reference model to exhaustively explore and predict the properties of very different types of assembled communities.

The picture that emerges is that complexity – understood as how difficult a phenomenon is to model – peaks at intermediate levels of heterogeneity. A bipartite network or a nested trophic hierarchy may cause deviations from the basic reference model, reflecting the need to add more structure (e.g. our

bioRxiv preprint doi: https://doi.org/10.1101/145862; this version posted June 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mixture model). Yet a highly multipartite network, such as a complicated web of functional groups and fluxes, would again tend to resemble a disordered set of species. Some types of structure, such as degree distributions, mutualistic network features, or consumer-resource mechanisms in competition, do not even seem to cause a significant deviation from generic collective behavior. Combining disorder and coherent structure to understand large systems has deep mathematical underpinnings: Terence Tao speaks of " a fundamental dichotomy between structure and randomness, which in turn leads (roughly speaking) to a decomposition of any object into a structured (low-complexity) component and a random (discorrelated) component." [42].

Future work should further ascertain when many model components can come together to create either order or disorder, and how our approach can be extended to many more settings by extracting the most essential structural information to complement the generic disordered dynamics. Finally, while we have focused on equilibrium properties, the methods we employed to analyze the reference model can be extended to its full dynamics [43], and hence, future investigations could provide an even broader understanding of generic dynamics in community ecology.

Acknowledgments

We thank Bart Haegeman for insightful discussions. This work was supported by the TULIP Laboratory of Excellence (ANR-10-LABX-41) and by the BIOSTASES Advanced Grant, funded by the European Research Council under the European Union's Horizon 2020 research and innovation programme (666971).

References

[1] Robert D Holt. Community modules. In *Multitrophic interactions in terrestrial ecosystems, 36th Symposium of the British Ecological Society*, pages 333–349. Blackwell Science Oxford, 1997.

[2] Jordi Bascompte and Daniel B Stouffer. The assembly and disassembly of ecological networks. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1524):1781–1787, 2009.

[3] Richard Levins. Discussion paper: the qualitative analysis of partially specified systems. *Annals of the New York Academy of Sciences*, 231(1):123–138, 1974.

[4] Ulrich Brose, Richard J Williams, and Neo D Martinez. Allometric scaling enhances stability in complex food webs. *Ecology letters*, 9(11):1228–1236, 2006.

[5] Anton Bovier. Statistical mechanics of disordered systems. Technical report, University of Aarhus. Centre for Mathematical Physics and Stochastics (MaPhySto)[MPS], 2001.

[6] Guy Bunin. Ecological communities with lotka-volterra dynamics. *Physical Review E*, 95(4):042414, 2017.

[7] Robert M May. Will a large complex system be stable? *Nature*, 238:413–414, 1972.

[8] Jacopo Grilli, Matteo Adorisio, Samir Suweis, György Barabás, Jayanth R. Banavar, Stefano Allesina, and Amos Maritan. Feasibility and coexistence of large ecological communities. *Nature Communications*, 8, 2017.

[9] Guy Bunin. Interaction patterns and diversity in assembled ecological communities. *arXiv preprint arXiv:1607.04734*, 2016.

[10] Edward H Simpson. Measurement of diversity. *Nature*, 1949.

[11] Ian Donohue, Helmut Hillebrand, José M Montoya, Owen L Petchey, Stuart L Pimm, Mike S Fowler, Kevin Healy, Andrew L Jackson, Miguel Lurgi, Deirdre McClean, et al. Navigating the complexity of ecological stability. *Ecology Letters*, 19(9):1172–1185, 2016.

bioRxiv preprint doi: https://doi.org/10.1101/145862; this version posted June 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[12] David Tilman. Biodiversity: population versus ecosystem stability. *Ecology*, 77(2):350–363, 1996.

[13] Russell Lande, Steinar Engen, and Bernt-Erik Saether. *Stochastic population dynamics in ecology and conservation*. Oxford University Press on Demand, 2003.

[14] Jean-François Arnoldi, Michel Loreau, and Bart Haegeman. Resilience, reactivity and variability: A mathematical comparison of ecological stability measures. *Journal of theoretical biology*, 389:47–59, 2016.

[15] Matthieu Barbier. Lecture notes: The cavity method for large ecosystem assembly. Unpublished paper, 2017.

[16] Robert Mac Arthur. Species packing, and what competition minimizes. *Proceedings of the National Academy of Sciences*, 64(4):1369–1371, 1969.

[17] Richard Levins and David Culver. Regional coexistence of species and competition between rare species. *Proceedings of the National Academy of Sciences*, 68(6):1246–1248, 1971.

[18] JE Cohen and CM Newman. A stochastic theory of community food webs: I. models and aggregated data. *Proceedings of the Royal Society of London B: Biological Sciences*, 224:421–448, 1985.

[19] J Nathaniel Holland, Donald L DeAngelis, and Judith L Bronstein. Population dynamics and mutualism: functional responses of benefits and costs. *The American Naturalist*, 159(3):231–244, 2002.

[20] J Nathaniel Holland and Donald L DeAngelis. A consumer–resource approach to the density-dependent population dynamics of mutualism. *Ecology*, 91(5):1286–1295, 2010.

[21] Vincent Calcagno, Nicolas Mouquet, Philippe Jarne, and P David. Coexistence in a metacommunity: the competition–colonization trade-off is not dead. *Ecology Letters*, 9(8):897–907, 2006.

[22] Sonia Kéfi, Vincent Miele, Evie A Wieters, Sergio A Navarrete, and Eric L Berlow. How structured is the entangled bank? the surprisingly simple organization of multiplex ecological networks leads to increased persistence and resilience. *PLoS Biol*, 14(8):e1002527, 2016.

[23] Robert H MacArthur and Edward O Wilson. *Theory of Island Biogeography*. Princeton University Press, 2015.

[24] Dominique Gravel, François Massol, Elsa Canard, David Mouillot, and Nicolas Mouquet. Trophic theory of island biogeography. *Ecology letters*, 14(10):1010–1016, 2011.

[25] David A Kessler and Nadav M Shnerb. Generalized model of island biodiversity. *Physical Review E*, 91(4):042705, 2015.

[26] Crawford Stanley Holling. The functional response of predators to prey density and its role in mimicry and population regulation. *Memoirs of the Entomological Society of Canada*, 97(S45):5–60, 1965.

[27] Simone Pigolotti, Roberto Benzi, Prasad Perlekar, Mogens Høgh Jensen, Federico Toschi, and David R Nelson. Growth, competition and cooperation in spatial population genetics. *Theoretical population biology*, 84:72–86, 2013.

[28] Jordi Bascompte. Disentangling the web of life. *Science*, 325(5939):416–419, 2009.

[29] Jordi Bascompte and Pedro Jordano. Plant-animal mutualistic networks: the architecture of biodiversity. *Annu. Rev. Ecol. Evol. Syst.*, 38:567–593, 2007.

[30] Paul Erdos and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[31] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

bioRxiv preprint doi: https://doi.org/10.1101/145862; this version posted June 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[32] Jose M Montoya and Ricard V Solé. Topological properties of food webs: from real data to community assembly models. *Oikos*, 102(3):614–622, 2003.

[33] Lai Zhang, Martin Hartvig, Kim Knudsen, and Ken H Andersen. Size-based predictions of food web patterns. *Theoretical ecology*, 7(1):23–33, 2014.

[34] M Gatto. A general minimum principle for competing populations: Some ecological and evolutionary consequences. *Theoretical Population Biology*, 37(3):369–388, 1990.

[35] Stefano Allesina and Mercedes Pascual. Food web models: a plea for groups. *Ecology letters*, 12(7):652–662, 2009.

[36] R. P. Rohr, S. Saavedra, and J. Bascompte. On the structural stability of mutualistic systems. *Science*, 345(6195):1253497–1253497, 2014.

[37] Jean-François Arnoldi, Bart Haegeman, Tomás Revilla, and Michel Loreau. Particularity of "universal resilience patterns in complex networks". *bioRxiv*, page 056218, 2016.

[38] Robert T Paine. Ecological determinism in the competition for space: The robert h. macarthur award lecture. *Ecology*, 65(5):1339–1348, 1984.

[39] Mikhail Tikhonov. Community-level cohesion without cooperation. *Elife*, 5:e15747, 2016.

[40] Jianxi Gao, Baruch Barzel, and Albert-László Barabási. Universal resilience patterns in complex networks. *Nature*, 530(7590):307–312, 2016.

[41] Chengyi Tu, Jacopo Grilli, and Samir Suweis. Collapse of resilience patterns in generalized lotka volterra dynamics and beyond. *arXiv preprint arXiv:1606.09630*, 2016.

[42] Terence Tao. The dichotomy between structure and randomness, arithmetic progressions, and the primes. *arXiv preprint math/0512114*, 2005.

[43] Yoshimi Yoshino, Tobias Galla, and Kei Tokita. Statistical mechanics and stability of a model eco-system. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(09):P09003, 2007.