Diffuse cliques maintain biodiversity in species-rich ecological communities

Matthieu Barbier1,3*, Claire de Mazancourt1, Michel Loreau1 and Guy Bunin2

May 28, 2019

High-dimensional phenomena, which often defy low-dimensional intuitions1,2, are an essential and yet seldom explored frontier in our understanding of ecological communities3. Ecologists have long speculated about how large numbers of species manage to coexist in rich assemblages. Most answers to date have focused on identifying particular dimensions along which species may organize to persist together4. Here we instead ask: what is the characteristic structure of a community where coexistence arises from a large number of concurrent factors? In such communities, individual species might not follow any evident pattern in their interactions, yet the group as a whole exhibits a statistical structure that we call “diffuse clique”. We find remarkable quantitative evidence for this pattern across a range of plant biodiversity experiments. Our approach exploits the emergent simplicity of high-dimensional systems5,6, a powerful idea originating in physics that has, so far, rarely been demonstrated unequivocally in ecological data. We conclude that a subtle form of collective order may underlie complex networks of species interactions. This diffuse order offers a new grasp on how ecological communities maintain their fascinating diversity.

The coexistence of many species with similar attributes is a long-standing puzzle: simple theories and experiments support the principle of competitive exclusion, whereby the best competitor should displace all others7,8. Yet, strict dominance by one species appears, at most spatial and temporal scales, to be the exception rather than the rule in the natural world. Over decades of ecological research, many partial solutions to this puzzle have been proposed, and integrated into the overarching framework of niche theory9. This framework suggests that we should identify particular trade-offs between abilities such as resource exploitation10, defense against predators11 and tolerance of temporal fluctuations12,13. Through these trade-offs, strict bounds are imposed upon how species grow and interact, preventing any species from overwhelming its competitors.

We propose to start from a different perspective. Coexistence in highly diverse communities likely involves a large number of niches and trade-offs, some known and many unknown a priori. Each cross-species interaction may be determined by a unique combination of factors, precluding any simple and conspicuous (low-dimensional) order in the community4. Some ecological theories therefore make the assumption that interactions are essentially random5 – a bold move, yet one that parallels major successes in physics14,15. Fully random interactions, however, do not allow many species to coexist4. The high biodiversity observed in many natural communities therefore implies some form of latent structure.

We first derive a theoretical prediction, the most parsimonious way to constrain species interactions in order to achieve coexistence. We uncover it by asking: if one samples many different interaction networks, and retains only those where all species survive, what do the remaining networks have in common? Some may appear very structured, others almost random. Yet, we find in Fig. 1 that most of these networks exhibit the same statistical pattern. This pattern, expressed in equations below, is a weak but crucial bias in how the most successful competitors interact with others16. We now derive this pattern from a simpler probabilistic argument, and explain in intuitive terms how it allows coexistence.

Measuring species interactions is often difficult and prone to high uncertainty17,18 and most empirical settings only give us access to aggregated statistics. The total effect of interactions on one species can be inferred from its relative yield

\[ \eta_i = B_i / K_i \]  

(1)

the ratio of its abundance \( B_i \) in a community to its abun-
dance without competitors $K_i$ (known as carrying capacity) in the same environment\cite{24}. We interpret species with higher $\eta$ as successful competitors, as they benefit more (or suffer less) in total from their interactions with others. The simplest way to model these interactions is by assuming a linear dependence between species

$$
\eta_i = 1 - \sum_{j \neq i} \beta_{ij} \eta_j \quad \text{for all } i.
$$

where $\beta_{ij}$ is the competitive effect of species $j$ on species $i$. This relationship, which can be tested empirically\cite{24,25}, holds between coexisting species at equilibrium in the classic Lotka-Volterra model.

Many different interaction networks can generate the same equilibrium community. Observing the coexistence of $S$ species with relative yields $\eta_i$ conveys some information about their interactions, but not enough to fully determine them: the equations (2) impose $S$ constraints, while there are $S(S-1)$ unknown interaction coefficients $\beta_{ij}$. On the other hand, community-wide statistics, such as the mean strength of competition $\bar{\beta}$, can be reliably deduced from that information\cite{29} (Appendix E).

We therefore adopt a probabilistic approach, and ask what is the most likely community structure, i.e. the set of features most widely shared among the many possible solutions. We first define a prior distribution $P(\beta_{ij})$ that can be adapted to our biological knowledge of the community. For all experiments below, we simply assume that each coefficient $\beta_{ij}$ is drawn independently from a normal distribution with mean $\bar{\beta}$. We then compute how this prior is modified once restricted to networks that admit the equilibrium $\eta_i$ (Appendix B). Computing a posterior distribution given a prior and linear constraints (2) is a well-established problem in probability theory\cite{26,27}.

We find that interactions $\beta_{ij}$ should follow two statistical patterns that both admit intuitive interpretations (Fig. 2). First, competition must be biased to explain which species are successful or not. If all interaction strengths were equal to the prior mean, $\beta_{ij} = \bar{\beta}$, we would expect any species to achieve the same relative yield,

$$
\eta^* = \frac{1 - \bar{\beta} \sum_i \eta_i}{1 - \bar{\beta}}.
$$

When $\eta_i > \eta^*$, we therefore expect that species $i$ suffers less competition than $\bar{\beta}$, and conversely if $\eta_i < \eta^*$. In our calculation, this appears in the conditional expectation of the competitive effect of $j$ on $i$,

$$
E[\beta_{ij}|\eta_i, \eta_j] = \bar{\beta} + (1 - \bar{\beta})\Delta(\eta_i, \eta_j)
$$

which deviates from the prior mean $\bar{\beta}$ by a bias

$$
\Delta(\eta_i, \eta_j) = \frac{(\eta_i - \eta^*) \eta_j}{\sum_{m \neq i} \eta_m^2}.
$$

Figure 1: Finding a general pattern of coexistence in species-rich interaction networks. Interaction networks of $S$ species are represented by $S \times S$ matrices (square boxes) where each element $\beta_{ij}$ denotes the effect of species $j$ on species $i$. Of all possible species interaction networks (bottom disk), only an infinitesimal fraction (shaded area) allows $S$ species to coexist at some equilibrium $\eta_i$. Zooming into this area of coexistence (upper disk), we find that most such networks appear almost random, yet they tend to follow a common trend which we call a “diffuse clique structure”: an underlying pattern of biases (4) and correlations (6) hidden in the large spread of coefficients $\beta_{ij}$. We define a metric $\rho$ to quantify how well our predicted pattern is observed in a given interaction network. We find in simulations that this metric $\rho$ gets closer to 1 as biodiversity increases (histograms show the distribution of $\rho$ for 200 networks with $S = 8$ and 25 species). By contrast, some coexistence mechanisms, such as the one-dimensional competition-colonization tradeoff (15,16), can give rise to highly atypical networks, showing unrelated or even opposite patterns.
We see that this bias is not evenly distributed. Competition coming from unsuccessful species (low $\eta_j$) can be random without compromising the equilibrium. On the other hand, a species that is successful (achieving high $\eta_j$) is likely to have biased interactions, competing less on average against other successful species, and experiencing weaker competition from them (Fig. 2).

The second pattern imposes that successful species $j$ and $k$ avoid competing against the same target $i$ (Fig. 2b).

$$\text{corr}(\beta_{ij}, \beta_{ik}|\eta_i, \eta_j, \eta_k) = -\frac{\eta_i \eta_j}{\sum_{m \neq i} \eta_m^2}. \quad (6)$$

While the first pattern determines the expected success of each species, the second pattern guarantees that each relative yield is exactly set to $\eta_i$. We show in Appendix B that this correlation pattern prevents $\eta_i$ from deviating from its expectation, which would likely drive some low-$\eta$ species to extinction in a fully random community.

Taken together, these two patterns suggest that we will generally observe a fuzzy “clique” of competitors that are both biased and successful, surrounded by unsuccessful species with arbitrary interactions. This picture differs in multiple respects from classic explanations of coexistence. It does not suppose a measurable segregation of species into distinct niches. By imposing only the weakest possible constraints upon the many degrees of freedom in $\beta_{ij}$, it allows interactions to take almost arbitrary values. It also represents a form of collective organization, where coexistence arises, not from particular species traits, but from statistical biases distributed over all interactions. Accordingly, Fig. 1 shows that this structure becomes increasingly prevalent (although more diffuse) in highly diverse communities.

We now present an empirical validation of these patterns on experimental data in Fig. 3 and Fig. 4. Grassland biodiversity experiments provide an ideal testbed for inferring species interactions and mechanisms of coexistence. Each experiment contains a large number of plots in which plant species are assembled in varying numbers and combinations, out of a pool of $S = 8$ to 60 species depending on the experiment. Biomass in monoculture (single-species plots) provides an estimate of the species’ carrying capacities.

To test our predictions, we split these data in two sets. Relative yields $\eta_i$ in the full-diversity plots are used to compute the theoretical expectations $\bar{\eta}$ and correlations $\text{corr}(\beta_{ij})$ of interactions. All other plots, comprising different subsets of the species pool, are used to fit individual interaction coefficients $\beta_{ij}$ by a multilinear regression of equation (2). From these fitted coefficients, we construct empirical estimates of the theoretical statistics.

We show in Fig. 3 the interaction matrix computed in the Wageningen grassland experiment. While lacking ap-
Using many species combinations (56 sets of S < 8 species), we can fit interaction coefficients \( \beta_{ij} \) by multilinear regression. Knowing only the relative yields \( \eta_i \) in the full community (all \( S = 8 \) species), our method suggests a theoretical expectation for each interaction, \( E[\beta_{ij}|\eta_i,\eta_j] \). Comparison between fitted coefficients and expectations. In our diffuse pattern, individual coefficients are expected to exhibit a large spread around their expectation \( E[\beta_{ij}] \). But we can construct an empirical estimate of the mean, the running average \( A[\beta_{ij}] \) (grey curve, 90% CI in shaded area). It is in good agreement with the theoretical mean, as shown by proximity to the red 1:1 line. We show this empirical average \( A[\beta_{ij}] \) in matrix form (median value for each species pair \( i,j \)), to compare with the predicted matrix \( E[\beta_{ij}] \) in (b).

The reconstructed statistical pattern in Fig. 3 agrees both qualitatively and quantitatively with the predictions. This striking agreement between theory and data is quantified for multiple experiments in Fig. 4. We stress that this is a strong test. All results are fully determined by measured abundances, without any adjustable parameter. Furthermore, none of the data used to parameterize theoretical formulas is involved in fitting the empirical interaction coefficients. Finally, we rule out these relationships being artefacts of our method, as they vanish for very sparse or noisy data, and can be violated in simulated ecosystems with a low-dimensional structure, as seen in Fig. 3 and 4.

The approach developed here provides a test of how typical an empirical or theoretical interaction network is, given the observed abundances of its species: how similar it is to the majority of possible networks admitting the same equilibrium \( \eta \) (Fig. 1). We also detail in Appendix B an algorithm for generating such typical networks. A deviation from typicality may hint at low dimensional mechanisms, such as particular trade-offs.

We have introduced a novel methodology for thinking about the collective organization of coexistence in ecological communities. This approach goes beyond the particular theoretical predictions (4) and (6), which are simplified results tied to our choice of unstructured prior distribution and linear interactions. When there is a positive but nonlinear relationship between data and predictions, our approach could be improved with more accurate inference and more realistic models, but it already captures an important qualitative feature of community organization. The same methods could be expanded by adding structure to the prior and nonlinearity to the dynamics. This will allow extensions to more complex communities, such as food webs, or networks that have been structured by other ecological and evolutionary processes. Future work should explore how this approach, based on ideas from statistical physics and generic properties of high-dimensional systems, can be generalized to other biological systems.
Methods

Experimental data

We employ data from 3 grassland biodiversity experiments in Wageningen, Netherlands\(^{20}\) and Cedar Creek, MN, USA (the Big Biodiversity\(^{31}\) and BioCON\(^{10}\) experiments). Each experiment uses a pool of species seeded or planted in various combinations, including some or all possible monocultures \((S = 1)\) species, some partial compositions, and all species planted together. We removed the first two year for all experiments as they showed clear evidence of transient dynamics (Appendix C).

Interactions measured in the Wageningen experiment showed much lower inference error than in other experiments. Therefore, we used this experiment to assess our hypothesis that observed abundances are primarily determined by fixed inter-species interactions (Appendix D). The consistency of the equilibrium Lotka-Volterra description \(^{2}\) was shown through a series of tests: we employed multiple inference procedures for the matrix \(\beta_{ij}\) and carrying capacities \(K_i\), using different subsets of the data for prediction and validation, and found them all statistically significant and in agreement within empirical uncertainty. In particular, carrying capacities \(K_i\) inferred from all multispecies plots \((S > 1)\) agreed with measurements in monocultures \((S = 1)\). Likewise, interactions \(\beta_{ij}\) inferred from all plots with \(S < 8\) were consistent with the equilibrium values of \(\eta_i = B_i/K_i\) in octoculture \((S = 8)\). This strongly supports the simple linear model \(^{2}\). Interaction estimates from other experiments were less robust and might be affected by nonlinearity, transient dynamics, stochasticity and errors.

Validation of theoretical predictions

For each experiment, we first split monoculture \((S = 1)\) replicates in two sets, and compute the species’ carrying capacities \(K_i\) for each set, to be used separately. For a maximal plot biodiversity \(S_{\text{max}}\), all plots with \(1 < S < S_{\text{max}}\) and the first set of monocultures were used to infer the interaction matrix \(\beta_{ij}\) using the hyperplane (multilinear) least-squares fit proposed by Xiao et al.\(^{32}\) (see Appendix D). The second set of monocultures and the species abundance \(B_i\) in plots with \(S = S_{\text{max}}\) were then used compute the relative yields \(\eta_i = B_i/K_i\) in the full community. All calculations were performed 250 times, using different bootstrapped sample means as values for \(K_i\) and \(B_i\). Each calculation led to a different set of \(\eta_i, \beta_{ij}\) and \(\beta\) (see Appendix A for calculation details).

We tested the two components of the diffuse clique pattern, starting with the pattern of means \(^{4}\). We plotted the measured values of \(\beta_{ij}\) (hereafter \(y\)) against their the-
oretical expectation \( E[\beta_{ij}|\eta_i, \eta_j] \) (hereafter \( x \)). To obtain an empirical estimate of the expectation for a single interaction coefficient, we performed a running average: for each point \((x, y)\), we replaced its \( y \) coordinate by the average \( \bar{y} \) within a window centered on \( x \) and spanning 10% of the \( x \)-axis; we also measure the 90% CI over bootstrapped values (Fig. 3). We then grouped all values \( \bar{y} \) associated with the same species pair \((i, j)\), took their median, and reconstructed an empirical matrix of expectations \( B_{ij} \) (shown in Fig. 4).

To construct a stringent test (see SI Appendix E), the metric of agreement \( \rho \) used in Fig. 1 and 4 is defined as the minimum of three correlation scores: \( \rho_0 \) testing the overall relationship between \( x \) and \( \bar{y} \); \( \rho_{row} \) testing the agreement of row-wise trends (within bins of \( \eta_i \)) and \( \rho_{col} \) for column-wise trends (within bins of \( \eta_j \)). This ensures that all the qualitative features described in Fig. 2 are present, and reduces the risk of spurious agreement scores.

We proceeded similarly to test the pattern of correlations \( \hat{C}_{ijk} \). Defining \( d_{ij} = \beta_{ij} - E[\beta_{ij}|\eta_i, \eta_j] \), we compute \( y = \delta_{jk} - d_{ij}/d_{jk}/\text{mean}(d_{mn}^2) \), where \( \delta_{jk} = 1 \) if \( j = k \) and 0 otherwise, and the denominator is the sample mean. We then plotted \( y \) against the prediction \( x = -\eta_i/\sum_{i \neq j} \eta_i^2 \), performed a running average to get \( \bar{y} \), and constructed an empirical tensor of correlations \( C_{ijk} \) from the median of all values associated with each species triplet \((i, j, k)\).

**Data availability**

This study brought together existing data that was obtained upon request (Wageningen biodiversity experiment data from J. van Ruijven\(^{22}\)) and data that is publicly available (Big Bio \( \text{http://www.cedar creek.umn.edu/research/data} \) and BioCON \( \text{http://www.biocon.umn.edu/} \)). Data represented in Fig. 3 and 4 is available at \( \text{https://github.com/mrcbarbier/diffuseclique} \).

**Code availability**

Computer code developed for this study is available at \( \text{https://github.com/mrcbarbier/diffuseclique} \).

**Acknowledgments**

We thank J.-F. Arnoldi, Y. Zelnik and F. Isbell for comments. Our gratitude goes to J. van Ruijven for sharing data from the Wageningen experiment. Experimental work at Cedar Creek was supported by grants from the US National Science Foundation Long-Term Ecological Research Program (LTER) including DEB-0620652 and DEB-1234162, and further support was provided by the Cedar Creek Ecosystem Science Reserve and the University of Minnesota. M.B., C.d.M. and M.L. were 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). G.B. was supported by the Israel Science Foundation (ISF) Grant no. 773/18.

**Author contributions**

G.B. developed the theoretical approach. M.B., G.B. and C.d.M. analyzed the data. M.B. wrote the first draft. All authors contributed substantially to study design and manuscript revisions.

**References**

[1] Jorge Kurchan and Laurent Laloux. Phase space geometry and slow dynamics. *Journal of Physics A: Mathematical and General*, 29(9):1929, 1996.

[2] Mikhail Tikhonov and Remi Monasson. Innovation rather than improvement: a solvable high-dimensional model highlights the limitations of scalar fitness. *Journal of Statistical Physics*, pages 1–31, 2018.

[3] James S Clark, Mike Dietze, Sukhendu Chakraborty, Pankaj K Agarwal, Ines Ibanez, Shannon LaDeau, and Mike Wolosin. Resolving the biodiversity paradox. *Ecology Letters*, 10(8):647–659, 2007.

[4] Adam Thomas Clark, Clarence Lehman, and David Tilman. Identifying mechanisms that structure ecological communities by snapping model parameters to empirically observed tradeoffs. *Ecology letters*, 21(4):494–505, 2018.

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

[6] Matthieu Barbier, Jean-François Arnoldi, Guy Bunin, and Michel Loreau. Generic assembly patterns in complex ecological communities. *Proceedings of the National Academy of Sciences*, 115(9):2156–2161, 2018.

[7] GF Gause. The struggle for existence. *Williams and Wilkins, Baltimore*, 1934.

[8] David Tilman. *Resource competition and community structure*. Princeton University Press, 1982.
