UNIVERSITÀ DEGLI STUDI DI PADOVA

Scuola di Dottorato di Ricerca in Fisica
XXVII Ciclo

RANDOMNESS AND CRITICALITY IN BIOLOGICAL INTERACTIONS

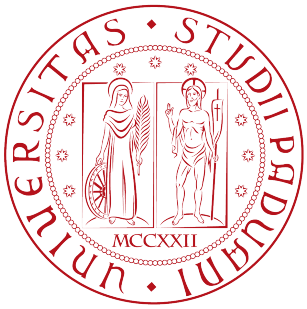
-

Direttore della Scuola: PROF. ANDREA VITTURI

Supervisore: PROF. AMOS MARITAN

Co-supervisore: DR. SAMIR SUWEIS

Candidato: JACOPO GRILLI



Dipartimento di Fisica e Astronomia 'G. Galilei'

ABSTRACT

In this thesis we study from a physics perspective two problems related to biological interactions. In the first part of this thesis we consider ecological interactions, that shape ecosystems and determine their fate, and their relation with stability of ecosystems. Using random matrix theory we are able to identify the key aspect, the order parameters, determining the stability of large ecosystems. We then consider the problem of determining the persistence of a population living in a randomly fragmented landscape. Using some techniques borrowed from random matrix theory applied to disordered systems, we are able to identify what are the key drivers of persistence. The second part of the thesis is devoted to the observation that many living systems seem to tune their interaction close to a critical point. We introduce a stochastic model, based on information theory, that predict the critical point as a natural outcome of a process of evolution or adaptation, without fine-tuning of parameters.

ASTRATTO

In questa tesi studiamo da una prospettiva fisica due problemi legati alle interazioni biologiche. Nella prima parte della tesi consideriamo le interazioni ecologiche, che danno forma agli ecosistemi e determinano la loro sorte, e la loro relazione con la stabilità degli stessi. Usando la teoria delle matrici aleatorie, siamo in grado di identificare gli aspetti chiave, i parametri d'ordine, che determinano la stabilità degli ecosistemi. Quindi consideriamo il problema di determinare la persistenza di una popolazione che vive in un territorio frammentato aleatoriamente. Usando alcune tecniche prese in prestito dalla teoria delle matrici aleatorie applicata ai sistemi disordinati, riusciamo a identificare quali sono gli ingredienti chiave per la persistenza. La seconda parte della tesi è dedicata all'osservazione che molti sistemi viventi sembrano essere calibrati precisamente vicino a un punto critico. Indroduciamo un modello stocastico, basato sulla teoria dell'informazione, che predice i punti critici come risultato naturale di un processo di evoluzione e adattamento, senza una calibrazione dei parametri.

PUBLICATIONS

Some ideas and figures have appeared in the following publications:

S. Allesina, J. Grilli, G. Barabas, S. Tang and A. Maritan. The Stability of Large Food Webs. Submitted to *Nature Communications*.
Section 2.7

S. Suweis, J. Grilli and A. Maritan. Disentangling the effect of hybrid interactions and of the constant effort hypothesis on ecological community stability. *Oikos*. 123(5):525-532. 2014.
Section 2.8

J. Grilli, M. Adorisio, S. Suweis and A. Maritan. Structural Stability in empirical and random mutualistic systems. *In preparation*.
Chapter 3

J. Grilli, G. Barabás and S. Allesina. Metapopulation persistence in random fragmented landscapes. Submitted to *Plos Computational Biology*.
Chapter 4

J. Hidalgo, J. Grilli, S. Suweis, M.A. Muñoz, J.R. Banavar and A. Maritan. Information-based fitness and the emergence of criticality in living systems. *Proceedings of the National Academy of Sciences*. 111(28):10095-10100. 2014.
Chapter 6

Other publications by the same author:

M. Adorisio, J. Grilli, S. Suweis, S. Azaele, J.R. Banavar and A. Maritan. Spatial maximum entropy modeling from presence/absence tropical forest data. *Preprint* on Arxiv: 1407.2425

J. Grilli, M. Romano, F. Bassetti and M. Cosentino Lagomarsino. Cross-species gene-family fluctuations reveal the dynamics of horizontal transfers. *Nucleic Acids Research*. 42(11):6850-6860. 2014.

J. Grilli, S. Suweis and A. Maritan. Growth or reproduction: emergence of an evolutionary optimal strategy. *Journal of Statistical Mechanics: Theory and Experiment*. 2013(10):P10020. 2013.

J. Grilli, S. Azaele, J.R. Banavar and A. Maritan. Absence of detailed balance in ecology. *Europhysics Letters*. 100:38002. 2012.

J. Grilli, S. Azaele, J.R. Banavar and A. Maritan. Spatial aggregation and the species-area relationship across scales. *Journal of Theoretical Biology*. 313:87-97. 2012.

L. Grassi, J. Grilli and M. Cosentino Lagomarsino. Large-scale dynamics of horizontal transfers. *Mobile Genetics Elements*. 2(3):163-167. 2012.

J. Grilli, B. Bassetti, S. Maslov and M. Cosentino Lagomarsino. Joint scaling laws in functional and evolutionary categories in prokaryotic genomes. *Nucleic Acids Research*. 40(2):530-540. 2012.

ACKNOWLEDGMENTS

I thought of how many places there are in the world that belong in this way to someone,
who has it in his blood beyond anyone else's understanding.

— Casare Pavese

Amos Maritan has been much more than a supervisor and a mentor in these years and I am thankful to him for a lot of reasons. From his curiosity and his desire to go beyond the borders of knowledge to his great lessons about small, yet very interesting, details on obscure statistical mechanics models. From the time spent in front of a blackboard to the belly laughs in his office. Above all, I want to thank Amos for the trust and the freedom he gave me. While he did that in a very natural way, it is something that is far from being obvious. Samir Suweis has being a constant presence during my graduate studies in Padova and I owe him a lot. I want to thank him for his sincere friendship and his ability to stay in front of the unknown and of the difficulties with a smile on his face. People in Padova have been great fellows in these years, and I want to thank all the people that populated my office (Denise and Laura expecially) and all the former, present and future members of Amos' lab: Claudio Borile, Marco Formentin, Milo Abolaffio (and his abdomen) and all the others. In particular, I am very thankful to *meu irmão* Rodrigo Pereira Rocha, for being so in love with life and for having taught me what *saudade* is, and to Matteo Adorisio, for his friendship, for being on the same wavelength on so many things (from *vincisgrassi* to *eleuthera*) and for reading this thesis.

Marco Cosentino Lagormasino has been the first person proposing me a research project, almost six years ago. I feel deeply thankful to him for many reasons. For the trust he put on me at the first time, for the trust he is still putting on me, for the way he collaborates with people and with me, for many advises he gave me during these years, for involving me in many projects and, most than everything, for the fun I have collaborating with him. I want to thank Bruno Bassetti for his unique, elegant, self-ironic and modest way of being a teacher, a researcher and a mentor. I wish also thank all the people in Marco's (enlarged) group I had the fun to collaborate with, and in particular Marco Gherardi and Matteo Osella.

Stefano Allesina welcomed me in his lab and I want to thank him for his enthusiasm and (together with Elena, Luca and Marta) for having let me feel at home while I was in Chicago. I wish also to thank all the members of Allesina's lab, and in particular Gyuri Barabás, for all the passionate discussions.

I am very thankful to the Andalusian collaborators. Miguel Muñoz for his enthusiasm and his *inteligencia emocional* and Jordi Hidalgo, for his great friendship and for the fun I always have when I work with him.

I am very grateful to all the people I had the pleasure to collaborate with (Sergei Maslov, Jayanth Banavar, Sandro Azaele and Federico Bassetti in particular) and to all the ones that I met at conferences and schools.

There are several people that are not directly involved in my research, but made it indirectly to be possible, fun and different. I wish to thank all the people I studied with in Milan, all my friends (especially all the green dots), my parents-in-law and siblings-in-law, and Aldo, who taught me how to love the universe behind any small object. I will be eternally grateful to my parents, Daniela and Emilio, to my sister Ilaria and to Filippo for the support and the patience. Last, but not least, to Sara, for nothing and for everything, that always coincide.

CONTENTS

0 INTRODUCTION

Philosophy is written in this grand book
– I mean the universe –
which stands continually open to our gaze,
but it cannot be understood unless one first learns to comprehend the
language in which it is written.
It is written in the language of mathematics.

— Galileo Galileo

0.1 A STOCHASTIC OVERTURE

This thesis is about interactions in biological systems. While we have a deep a understanding and an elegant theory (or perhaps elegant theories) on the forces that shape the universe, from quarks to galaxies, we still lack a comprehensive view of how interactions between molecules make things to be alive.

We do not want to go into definitions and face the problem of defining life [2]. Life. There is something clearly different in life, as we immediately recognize the difference between what is animate and what inanimate. We do not believe that a "life force" able to animate the inanimate is the explanation. As physicists studying life as a phenomenon, we are trying to understand how the microscopic forces between inanimate objects generate life as an emergent property.

Physicists, as natural scientists, as always been interested in the phenomenon of life and in its quantitative description. Helmholtz, Maxwell, Rayleigh did not raise barriers on their interests and seemed less interested in what we are on distinguish physics, chemistry or biology. Delbrück is the shining example of how much physicists contributed to biology.

We are pretty far from having any elegant, concise and comprehensive theory of life. Yet, we have principles, that sustains and guide hypothesis [3]. *Randomness is important*. From DNA trascription to population dynamics noise plays an important role, is the driving source of mutations and variability of the environment force living organisms to be flexible. *No fine-tuning*. Life is a robust phenomenon in a noisy and variable world. *Information and representation*. Survive and procreate is always about extract information from an environment, represent it and pass the information to others individuals or next generations.

0.2 WHAT IS IN THIS THESIS

This thesis is about randomness, fine-tuning and information passing in biological interactions and it is divided in two self-consistent part.

In the first part we consider explicitly one specific class of interactions, ecological interactions and we will try to make a connection between interactions and stability of ecosystems. Chapter 1 will define what are ecological interactions and their different types. We will describe them as networks identifying different topological structures that characterize empirical ecological interactions and we finally introduce the problem of stability of ecosystems. In chapter 2 we describe an approach to study the relation between interactions and stability, based on random matrices. We will also present original results about the stability of large structured food-webs and the stability in presence of different interaction types. Chapter 3 is completely original and is about mutualistic systems and the size of stability domain. In this chapter we introduce a numerical way to compute how big is the domain of conditions that leads to stable ecosystems and we put in relation with the structure of empirical networks. In chapter 4 we do not consider interactions between species but interactions between spatial locations. We introduce an original way to estimate the persistence of a population in a randomly fragmented landscape base on random matrices.

If the first part was mainly about noise in interactions, as interactions were modeled using random matrices, the second part is about fine-tuning of interactions and their role in information passing. Chapter 5 will propose some evidences that living systems operate close to a critical point. We will also suggest some possible biases that could affect this observation. In chapter 6 we propose an original model trying to suggest a mechanism that connects criticality with information passing, showing how the former can emerge without fine-tuning of parameters.

Part I

STABILITY AND RANDOMNESS IN ECOLOGICAL INTERACTIONS

Perhaps randomness is not merely an adequate description for complex causes that we cannot specify. Perhaps the world really works this way, and many events are uncaused in any conventional sense of the word. Perhaps our gut feeling that it cannot be so reflects only our hopes and prejudices, our desperate striving to make sense of a complex and confusing world, and not the ways of nature.

— Stephen J. Gould

1 ECOLOGICAL INTERACTIONS AND STABILITY OF ECOSYSTEMS

The first law of ecology is that
everything is related to everything else.

— Barry Commoner

1.1 ECOLOGICAL INTERACTIONS

Individuals living in an ecosystems are not independent entities. From microscopic viruses to gigantic plants, from bacteria to mammals, every living form interacts with the environment and with other living forms. Viruses need microbes or other cells to reproduce, plants need bacteria in the soil to grow and mammals have to eat plants or other animals to survive. An ecosystem is therefore not just a sum of individuals and species, but also a set of interactions and interdependences.

As we already mentioned in the introduction of this thesis, interactions in biology and in ecology have not (yet) the simple and elegant form they have in the physics of inanimate matter. Interactions between individuals and species depend, at least as far as we know, on the peculiar characteristics of the organisms we are considering, on their evolutionary history and on all the possible contingencies that affect their life and their evolution. It could seem therefore hopeless to study this kind of systems from any general perspective and that the only description possible is an enumerations of differences. As we show in sections 1.2 and 1.3 this is fortunately not the case. When interactions are described in terms of networks, striking regularities emerge, revealing the universal properties that shape ecological interactions.

Ecological interactions define the interdependence between species or groups of individuals. In section 1.4 we introduce a dynamical description of species, where populations are coupled by interactions. One can therefore study the dynamical properties of ecosystems when they are seen as dynamical systems. As one is not typically interested to transients, the stationary properties are the main focus of research. The natural question that has to be asked when dealing with stationary properties is whether they are stable or not. Since we are dealing with ecosystems determine the stability is very relevant, as it coincides with the theoretical side of the problem of conservation. In section 1.5 we introduce the problem of stability in ecology, that will be fully explained and described in chapter 2.

1.2 INTERACTIONS AS NETWORKS

Complex systems interact in a complex way. The standard way to look at interactions in a complex world is to look at them as networks.

A network (or a graph) is a collection of N nodes and L edges (or links), where pairs of nodes are connected by edges. A social network is just an example, where people (nodes) are connected by friendship (edges). Social networks are an example of an undirected network, where edges have not an orientation (friendship is a mutual relation). If edges have orientation, e.g. in the world wide web, where nodes represent Web pages and edges are the hyperlinks between pages, the networks is directed.

A network of N nodes can be represented as a $N \times N$ matrix, called adjacency matrix. Each element of the matrix represents the presence or the absence of a connection between nodes. If A is the adjacency matrix, the element A_{ij} is equal to zero if there is not a connection from node i to j, while it is equal to 1 if there is a link between them. If the network is undirected the adjacency matrix is symmetric, as a link between node i and node j always corresponds to a link between j and i.

Properties of networks have been studied from several different point of views. Physics, and in particular statistical mechanics, have contributed a lot in recent years. These studies has mainly focused on finding non-null regularities in empirical networks and explain them using simple models. Regularities has been found across several different empirical networks and what clearly emerges is that their are *complex*, where *complex* means that their structure and their topological properties are non-trivial being very different from what one would expect for simple graph.

The definition of *complex* relies then on the definition of *simple*. Example of simple graphs are regular graphs, where each node have exactly the same properties than any other (e.g. a d-dimensional lattice), and Erdös-Rényi (ER) random graphs, that are built by assigning a link between two nodes independently and with equal probability. In these simple networks all nodes have the same or similar degree (the number of links starting from each node, e.g. the number of friends if we are considering a social network) and the distribution of degrees (i.e. the probability of finding a node with a given degree) is a well behaved distribution peaked around a value (e.g. a Poisson distribution in the case of ER random graphs). Simple networks are well diluted, meaning that different groups of nodes have similar properties. example. The number of friends each person have is strongly related to its social status, varying a lot between people [5] and people having a common friend are more likely to be friend than what expected by random [6]. Example of properties that identify a network as complex are: scale-free degree distribution (the degree distribution has

a power-law tail), modularity (groups of nodes that have more connections between nodes of the same group than between nodes of different groups), assortativity or dissortativity (nodes with high degree tend to have more connections with nodes of high/low degree) and many others. Going back to our example, social networks show a power-law degree distribution [5], are modular [6] and assortative [7].

Ecological interactions can be represented as networks. Species are nodes and a link between two nodes represents an interaction. As there are different types of ecological interactions there are different types of interaction networks. In the next section we will introduce different interaction types and their properties in terms of networks.

1.3 STRUCTURE OF ECOLOGICAL INTERACTIONS

Ecological interactions are not of the same type. Both lions and gazelles and bees and flowers interact, but they do that in very different ways. There are three main types of interactions, depending on the effect they have. If the interaction is beneficial for both the individuals, it is said to be *mutualistic* (e.g. between bees and flowers), while if the interaction is disadvantageous for both, then it is *competitive* (e.g. between hyenas and lions competing for a prey). If the the interaction is beneficial for one and disadvantageous for the other, the interaction is said to be *antagonistic* and can be either *predation* (e.g. lions eating gazelles) or *parasitism* (e.g. viruses and bacteria). As will be explained in section 1.4, being advantageous or disadvantageous is related to an increasing or a decreasing of the population abundance of the individuals that are interacting.

Observational data gives the possibility to build an interaction network between species describing their interactions. The structure and the properties of these empirical networks are the focus of this section. We analyze in detail two types of networks corresponding to two interaction type: mutualistic networks and predator-prey networks (food-webs). There will be also discussed possible models that try to reproduce empirical networks.

1.3.1 *Mutualistic networks*

Empirical mutualistic networks are typically undirected bipartite networks. The nodes of a bipartite network can be divided in two disjointed sets, such that every edge connect one node of a set to one node of the other, and two nodes of the same set are never linked. A bipartite graph can be equivalently defined as graph that does not contain any cycle of odd length.

The empirical networks available are typically plant-pollinator networks or plant-seed disperser networks. In the former case the two sets of the bipartite mutualistic network are species of plants and

species of pollinators, where a link between them indicates a mutualistic interaction (plants have an advantage as the disperse their pollen, while pollinators get food). In a very similar way, in plants-seed disperser networks, plants get their seed dispersed by animals that eat their fruits.

Mutualistic networks do not have a random structure [8, 9]. Their degree distribution is highly heterogeneous: there are specialistic species (interacting with few other species) and generalist ones (that seem to interact with many others). Moreover these networks are said to be nested, with specialists interacting more likely with generalists than with other specialist [8]. This statement, "specialists interacting more likely with generalists", is a relative statement, meaning that "more likely" refers to what we should expect by chance. Defining the null expectation, i.e. what we should expect by chance, requires the definition of a null model. Even more trivially we have to define a measure of what nestedness is. Despite there is a general consensus that mutualistic network are nested, there is still a debate around these two problems: how to measure nestedness and what is the right null model. The commonly accepted measure is NODF [10], even though it presents several issues. A simpler version of it, known as η, is also used. Recently it as been propose to use the spectral radius of the adjacency matrix [11].

The debate around the correct null model to use is probably more subtle. The simplest null model consists in rewiring the links, conserving therefore only the dimensionality of the network and its number of links, constraining the randomized network to be fully connected. This randomization corresponds to generate Erdös-Rényi random graphs with a given number of links, and it usually called Erdös-Rényi (ER) randomization. All the measures of nestedness have values significantly much larger in empirical networks than in their ER randomizations. Figure 1 shows the empirical nestedness (NODF) measured in emprical netwroks compared with the distribution of values found out of ER randomization. Except very few cases, all the networks show a NODF larger than what expected by chance.

This fact could be a consequence of the heterogeneity of degrees. Therefore, nestedness has been studied in relation to the degree distribution of empirical network. Typically degrees are conserved "on average" in randomizations [12], meaning that a connection between i and j is considered with probability proportional to $k_i k_j$ (where k_i is the empirically observed degree of the node i). Again these randomizations show a tendency of networks to be nested [12]. This type of randomization has been criticized, since it biased the result toward more dissortative network [13, 12]. The most conservative randomization conserves exactly the degree sequence (i.e. the observed k_i). Being a very conservative randomization it produces very few networks, and there are in principle cases where there is only one graph

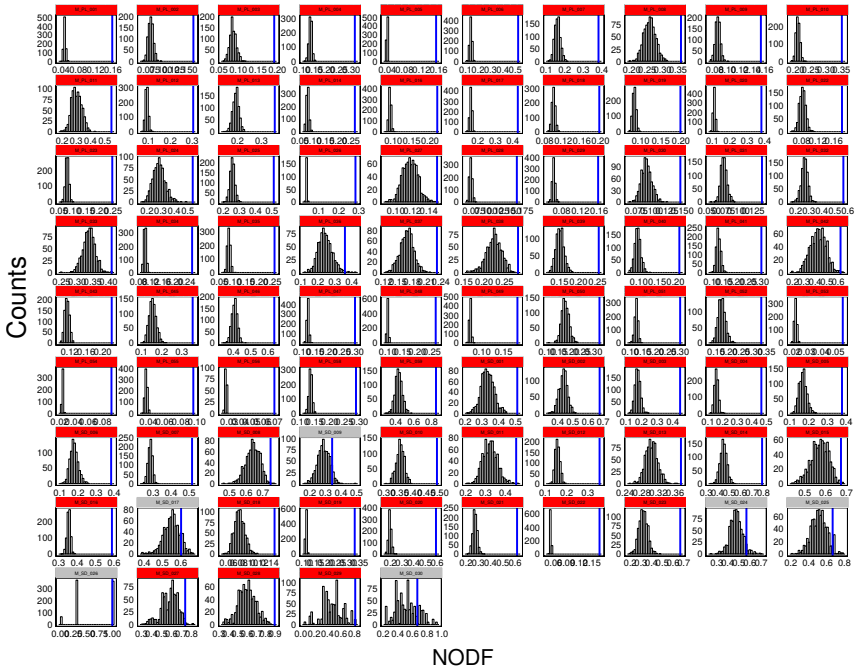


Figure 1: Nestedness in 85 empirical mutualistic networks. Each panel refers to a network (either a plant-pollinator if the name starts with M_PL or plant-seed disperser if the name starts with M_SD). The x-axis represents NODF [10] and the blue line is the value measured in the empirical network. The black bars are the histogram obtained by measuring the NODF over the ER randomizations. Panels with the red label indicate an empirical value of NODF significantly larger than randomizations (p−value smaller than 0.025), while gray ones mean that the results are not significant. Empirical networks are significantly more nested than what expected by chance, when only the number of links is conserved.

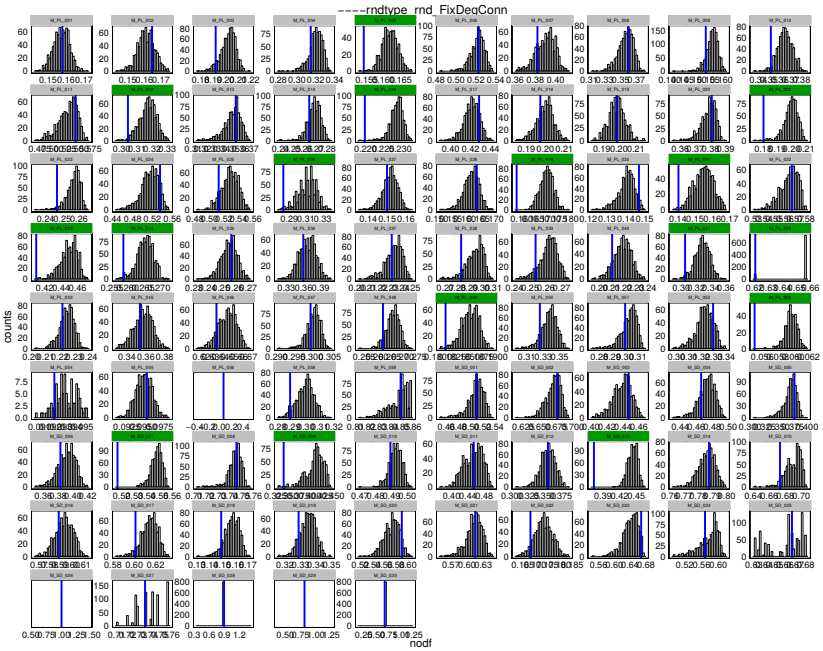


Figure 2: Nestedness in 85 empirical mutualistic networks. Each panel refers to a network (either a plant-pollinator if the name starts with M_PL or plant-seed disperser if the name starts with M_SD). The x-axis represents NODF [10] and the blue line is the value measured in the empirical network. The black bars are the histogram obtained by measuring the NODF over the randomizations that conserve the exact degree sequence. Panels with the green label indicate an empirical value of NODF significantly lower than randomizations (p−value smaller than 0.025), while gray ones mean that the result are not significant. network not significantly more nested than what expected by chance, when the degree sequence in conserved.

having the observed degree sequence [11]. Figure 2 shows that empirical networks are not more nested than networks with the same degree sequence.

A lot of speculations have been made on why nestedness could be beneficial for empirical networks [14], and different mechanisms leading to nested graph have been [15, 16, 17, 18] proposed. In particular nestedness has been showed to emerge from an adaptive/evolution dynamics in which species rewired their links in order to maximize their population abundance [18].

1.3.2 *Food-webs*

Food-webs represent the connections between species via the food chain. They are directed networks, where a link from species i to species j indicates that i predates j (sometimes the opposite notation

is used, where a link indicate the flow of energy/biomass from an organism to the other).

Despite the first food-web was graphically represented by Lorenzo Camerano in 1880, they have been studied quantitatively as networks only recently, with the goal of identifying common properties rather than peculiarities. Many properties that classify them as complex networks have been identified, such as non-poisson degree distribution (even if it is not power-law [19]), non-null network motifs [20], showing some universal properties [21].

Since the big fish (usually) eats the small(er) fish, food-webs are more directional than random directed networks. They contains few directed loops [21] and species tend to be ordered in levels (trophic levels), with incoming links (predators) from upper levels outgoing links to lower levels (preys). A simple quantitative model that considers directionality is called cascade model [22]. In the cascade model, S species are ranked and for each species j, a link from one of the species with lower rank is drawn with probability C. If K is the adjacency matrix, when $C = 1$, the upper-triangular part of K is completely filled with ones. The cascade model produces directed acyclic graphs, so that all the non-zero coefficients of K are contained in its upper-triangular part.

The cascade model considers the directionality of empirical food-webs, but ignore one other important ingredient that is intervallity. If it is true that the big fish eats the small fish, it is also true that sharks do not eat shrimps. Most of the species (with some exceptions) tend to predate species that are smaller but not too much. The niche model [23] takes inspiration from this fact. Building an adjacency matrix using the niche model requires four steps:

1. Each species j is assigned a "niche value", η_j, sampled from a uniform distribution $\mathcal{U}[0, 1]$, and the species are sorted in increasing order.

2. A "niche radius" is obtained for each species: $r_j = \eta_j B$, where B is a value sampled from the beta distribution $\mathcal{B}(1, \beta)$.

3. A "niche center", c_j, is sampled from $\mathcal{U}[r_j/2, \min(\eta_j, 1 - r_j/2)]$.

4. Species j consumes all species k whose niche value is included in an interval: $(c_j - r_j/2) \leqslant \eta_k \leqslant (c_j + r_j/2)$.

To ensure that we obtain a number of links that is about right [24], one has to set $\beta = S(S-1)/L - 1$. The niche model well reproduces the main features of empirical food-webs [23], being also very parsimonious in parametrization (as depends only on the number of species and the number of links).

1.4 DESCRIBING THE INTERACTIONS IN AN ECOSYSTEM

One can think of an ecosystem as a stochastic dynamical system. The populations of species change in time because of stochasticity of the environment, demographic processes and the fact that different populations are coupled via interactions. In this section we will not introduce stochastic dynamics, and we will describe the dynamics of an ecosystem as it was purely deterministic. The equations that follow can therefore be interpreted as an appropriate deterministic limit of stochastic equations or an equation for the average population under a proper mean-field approximation.

The fundamental equation of population growth of a single species is the logistic equation

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) , \tag{1}$$

that models the growth of a population in presence of limiting resources. The parameter K is known as carrying capacity, and represents the maximum population size that the resources can sustain, while r is the per-capita growth rate. If the resources were infinite ($K \to \infty$), the population would grow exponentially, with rate equal to r. The resulting dynamics in presence of limiting resources is quite simple, if $r > 0$, the population converges to a globally stable equilibrium $N^* = K$, while if $r < 0$, the stable equilibrium is $N^* = 0$.

The logistic equation, even if it considers only one species, already contains some ecological information. The self-limitation term can indeed be considered as the result of a competitive interaction between individuals of the population. In a word with infinite resources there is no competition, and every individual would have the same probability to die/have progeny independently of the size of the population. When the resources are limited, the more the individuals are the less the resources per individual are, giving rise to a decreasing of the growth.

The most celebrated equation describing multiple species is probably the Lotka-Volterra equation. It describes the growth of two populations, a predator and a prey. Instead of writing it in its original formulation, we consider its generalization to multiple species and multiple interaction types

$$\frac{dn_i}{dt} = r_i n_i + n_i \sum_{j=1}^{S} A_{ij} H_i(n_j) , \tag{2}$$

where n_i is the population of species i. The matrix A is called interaction matrix. Its element A_{ij} describes the effect of the population abundance of one species j on the species i, and their sign completely identify the interaction type. If both A_{ij} and A_{ji} are positive the interaction is mutualistic, if both are negative the interaction is competitive, while if one is positive and the other one is negative we are

dealing with an antagonistic interaction. The diagonal elements are usually negative and represent the competition between individuals of the same species. The function $H_i(\cdot)$ is a positive function and is called *functional response*. Suppose that we are dealing with predator-prey interaction. The functional response quantifies how much the per-capita growth rate of a predator changes as we change the population abundance of preys. We can expect in principle this response not to be always linear: if the number of preys is much larger than the number the predators can handle, further increasing it does not have any effect on the population of predators. This function is modeled in several different ways. The simplest case is a linear functional response $H_i(n_j) = n_j$, that does not take into account the expected saturation. This is also called Holling type-1 functional response. The most used functional form is called Holling type-2, where

$$H_i(n_j) = \frac{n_j}{1 + h_i n_j} \,, \tag{3}$$

where the parameters h_i are called *handling times*, and can be interpreted as the time needed for a predator to handle a prey. The number of possible complications and details that one can introduce is clearly very large. More the complications are more is difficult to deal with the dynamical properties of the model. In the next section we will introduce the problem of stability of these dynamical systems.

1.5 STABILITY OF AN ECOSYSTEM

The ecosystems around us seems to be fairly stable, at least on our time scales. One would like to know how much they are stable and how the stability depends on the interactions between species. The word *stability* refers always to a stationary property and to a type of perturbation. In this thesis we will refer to stationary fixed point and to small perturbations on it in chapters 2 and 4, while in chapter 3 we will explore the effect of perturbations on parameters. In the first case stability means local asymptotically stability, i.e. the ability of a stationary fixed point to return to its equilibrium value after a small perturbation.

The hypothesis of fixed points and small perturbations seems to be a brutal approximation, even more brutal if we use deterministic and not stochastic equations. The deterministic hypothesis can be justified by the fact that for any stochastic dynamics we can write deterministic equations for their moments (see Appendix A). The equations for the average populations will depend on their variances and correlations, but one can always write enough complicated close equations involving only the averages (mean-field equations being the most brutal approximation). Any dynamics results in one of three possible stationary outcome: fixed-points, limit-cycles or chaos. Despite it is difficult to detect it, chaotic behaviors have almost never been identified

in empirical population time-series. Limit cycles and fixed points are completely different from a dynamical perspective, but in empirical time series (where fluctuations play a role) do not look very different. It is not that brutal to consider only fixed points and stability under small perturbation is the necessary condition for stability under largest ones.

At this point one could imagine that in order to study the stability of an ecosystem, one would need to choose a set of dynamical equations, complicated enough to be 'realistic', but not too much, as they must be tractable, and then study the dynamical properties of the system. In the next chapter we will introduce a different approach to the problem of stability based on random matrix theory. This approach does is not assume anything on the underlying dynamics and it is based on parameterizing interactions and properties closed to a fixed point as random variables, allowing us to identify few key 'order' parameters that determine the stability of an ecosystem.

2 RANDOM MATRICES AND STABILITY

The Edge...
There is no honest way to explain it
because the only people who really know where it is
are the ones who have gone over.

— Hunter S. Thompson

2.1 WILL A LARGE COMPLEX SYSTEM BE STABLE?

In principle, to study the stability of an ecosystem one would have to consider its dynamics in fully generality. This task becomes very difficult as the system becomes sufficiently large and as one introduces many details in the dynamics. Moreover it would require the full specification of all the parameters defining the dynamics.

In his seminal paper of 1972 [25], Robert May introduced with simple arguments a new way to look at stability in ecological systems. The original idea was based on considering an arbitrary dynamics describing the population dynamics of an ecosystem

$$\frac{d\underline{n}}{dt} = \underline{f}(\underline{n}) \,, \tag{4}$$

Let assume the existence of a stationary point $\underline{n}^*$ (defined as $\underline{f}(\underline{n}*) = 0$) and linearize the dynamics around it

$$\frac{d\delta x_i}{dt} = \sum_j A_{ij}\delta x_j \,, \tag{5}$$

where $\delta x_i = n_i - n_i^*$ and the matrix A is the Jacobian evaluated at the equilibrium point

$$A_{ij} = \frac{df_i(\underline{n})}{dn_j}\Big|_{\underline{n}=\underline{n}^*} \,. \tag{6}$$

The matrix A contains all the information about the dynamics as its state is sufficiently close to the stationary one. In particular it describes the stability of the stationary point $\underline{n}^*$, being the latter stable if all the eigenvalues of A have negative real part. The elements of A of course depend on the dynamics and the stationary point (i.e. the function $\underline{f}(\underline{n})$). This matrix is known in ecology as *community matrix*.

May's original idea was to model A as a random matrix. In particular one can imagine that it reflects the interactions between species. The diagonal elements of A represents a signature of intra-specific interactions (e.g. self-limitation), and therefore their average was fixed

equal to -1 without loss of generality. The matrix $B = A + 1$ is a random matrix, whose elements have mean equal to zero. With probability $1 - C$ an element of B is equal to zero, while with probability C it is drawn from a probability distribution with null mean and variance σ^2 The only assumption on the probability distribution is to have a finite fourth moment. Under these assumptions one can obtain a criterion for the stability of the system, by studying the distribution of eigenvalues of the random matrix and in particular the eigenvalue of the largest real part.

In particular (see section 2.4) one can prove that, if the probability distribution is a Gaussian and $C = 1$, the eigenvalues of B are distributed uniformly inside a circle in the complex plane of radius $\sigma\sqrt{S}$ (Girko's law). Due to universality (see section 2.2) this distribution does not depend on the choice of the distribution of matrix elements (if few mild assumptions on higher moments are satisfied [26, 27]) and we can therefore easily consider the case $C = 1$ by noticing that the variance of the matrix elements is equal to $C\sigma^2$. Therefore, since $A = B - 1$, the largest real part of the eigenvalues is, in the limit of large S, equal to

$$\langle \lambda_r \rangle = \sigma\sqrt{CS} - 1 \,, \tag{7}$$

where the average is taken over the randomness of the matrices. Nevertheless in this case the average value is also the typical one. In particular one can compute the probability of being unstable, and it turns out to be, in the large S limit, $P \to 0$ if $\sigma\sqrt{CS} < 1$ and $P \to 1$ if $\sigma\sqrt{CS} > 1$. The transition between stability and instability is very sharp. In particular its relative width scales as $1/S^{2/3}$ [28].

This result has very important consequences. Firstly it predicts a sharp transition between stability and instability governed by few parameters (in this case S, C and σ). Therefore it clearly indicates what are the relevant parameters that control stability and what is their relative contribution. Secondly, it predicts that largest and more connected systems tend to be less stable. This point has been considered by many ecologist as a paradox, since one would expect that biodiversity increase stability, and have triggered a huge debate that is not yet resolved.

In the next sections we review some results obtained in the context of random matrix theory and their application to stability of dynamical systems (with a particular focus on ecology).

2.2 RANDOM MATRIX ENSEMBLES

Random matrices are matrices whose elements are randomly drawn from a distribution $P(M)$. The main object of research in random ma-

trix theory is the distribution of eigenvalues (also known as density of states), which is defined by

$$p(\lambda) = \frac{1}{N}\sum_{i=0}^{N}\langle\delta\left(\lambda-\lambda_i^M\right)\rangle\,, \tag{8}$$

where λ_i^M is an eigenvalue of a $N\times N$ matrix M and the average is taken over the random matrices. There main goal is typically to find the distribution $p(\lambda)$ in the large N limit. In appendix B it will be shown an example of such a calculation.

Interestingly there is an increasing evidence that many properties of random matrices (e.g. the distribution of eigenvalues) are universal. Universality in this context means that those properties do not depends on the full distribution of $P(M)$ but only on few moments of the distribution.

2.3 SEMICIRCULAR LAW

The simplest and most known random matrix ensemble is the Gaussian orthogonal ensemble. It consists of symmetric matrices whose elements are drawn from a Gaussian distribution, with zero mean and variance equal to σ^2 (typically is considered, without loss of generality, a variance equal to $1/S$).

$$P(M) = \frac{1}{Z}\exp(-\frac{S}{4}\operatorname{tr}(M^2))\,. \tag{9}$$

The density of eigenvalues was firstly obtained by Wigner. It converges, in the large S limit for $\sigma^2 = 1/S$, to Wigner's semicircular law

$$p(\lambda) = \frac{2}{\pi}\sqrt{1-\lambda^2}\theta(1-\lambda^2)\,. \tag{10}$$

This results was firstly obtained by Wigner by solving all the moments of the distribution and can be obtained with several methods. Notably, in this case it is also possible to obtain [29], not just the distribution at large S, but the exact distribution of eigenvalues at any S. In appendix B we show an example of calculation leading to this solution (see appendix B).

The semicircular law plays a crucial role in random matrix theory. Matrices are peculiar random variables since they are non commuting objects. Independence for commuting variables implies lack of correlation. It is simple to realize that correlation between non-commuting random variables is an odd-defined quantity. This fact is a simple expression of a much more fundamental difference between commuting and non-commuting random variables, that requires new concepts in probability theory. In particular a new definition of independence need to be introduced for non-commuting random variables (known as *free probability*). *Free probability* [30] is the theory that

studies non-commuting random variables, and *free independence* is the concept generalizing independence in ordinary variables. In this context semicircular law plays a crucial role, specifically it plays the very same role that the Gaussian distribution plays in ordinary probability theory. The Gaussian distribution is the stable distribution respect to convolution, the distribution of the sum of independent random variables tends to a Gaussian, it is the distribution that maximizes the entropy constrained to a mean and variance, all the cumulants of degree larger than two are zero if and only if they are Gaussian distributed. In free probability theory, one can naturally introduce the concept of free cumulant, free convolution and free entropy. Using these concepts, the semicircular law is the stable distribution respect to free convolution, the distribution of the sum of free independent symmetric random matrices tends to a semicircular law, it is the distribution that maximizes the free entropy constrained to a mean and variance of the eigenvalue distribution, all the free cumulants of degree larger than two are zero if and only if they are distributed as a semicircular law.

For symmetric Gaussian matrices it is possible to calculate not only the distribution of eigenvalues, but also the probability distribution of the largest eigenvalue of a matrix. It is distributed accordingly to the Tracy-Widom distribution (which was introduced exactly to describe this quantity).

2.4 NON-SYMMETRIC RANDOM MATRICES

The results on non-symmetric (or non-hermitian) random matrices are much more difficult to obtain. A simple case that can be solved consists in random matrices with Gaussian entries and arbitrary correlation between off diagonal elements. In this case the distribution describing this ensemble is

$$P(M) = \frac{1}{Z} \exp(-\frac{N}{2(1-\tau^2)} \operatorname{tr}(MM^t - \tau M^2)) \, . \tag{11}$$

M^t is the matrix transpose of M, while τ is the correlation between elements opposed to the diagonal (i.e. M_{ij} and M_{ji}).

In this case the eigenvalues are uniformly distributed in an ellipse in the complex plane (see Figure 3). The semi-axes of the ellipse are equal to $1+\tau$ (real axes) and $1-\tau$ (imaginary axes). In the case of $\tau = 0$ (absence of correlation between all the elements) the distribution is uniform in a circle, which is known as Girko's Law [31].

2.5 EFFECT OF DIFFERENT INTERACTION TYPES ON STABILITY

As explained in chapter 1 species interact in an ecosystem in many different ways. Mutualistic interactions are positive for both the inter-

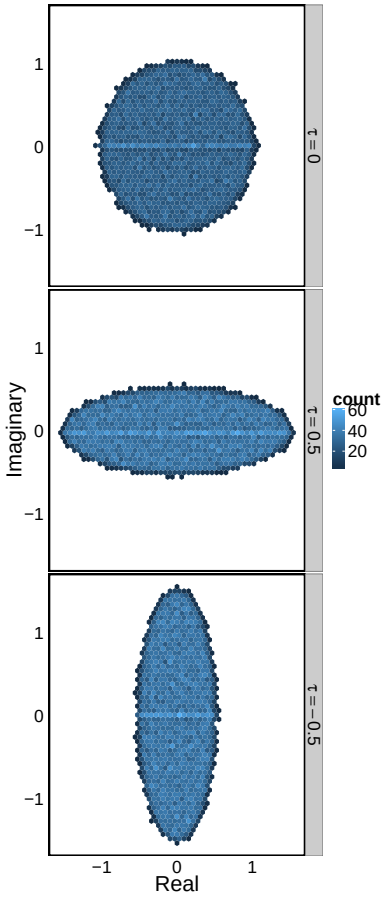


Figure 3: Distribution of eigenvalues in the complex plane. Random matrices were drawn from the distribution of equation 11, with different values of τ. The distribution of eigenvalues is uniform in an ellipse as predicted by [28], whose semi-axis depend on τ. The case $\tau = 0$ correspond to a circle. One can note that there is a deviation from the uniform density on the real axes, where there are on average more eigenvalues. As shown in [28], the fraction of real eigenvalues decrease as a power-law of the size, being an observable effect also at relatively large size. This figure was obtained by averaging over 50 matrices 500×500.

acting species, competitive interactions are disadvantageous for both while predation and parasitism favors one and disfavors the other one. These sign of interactions reflect to the way populations (and their fluctuations) are dynamically coupled. In particular, using the same approach explained in section 2.1, one can imagine a dynamics linearized around a proper stationary point. The linearization matrix A will reflect the sign of the interactions between the corresponding species.

The result obtained by May is the simplest possible scenario, in which the interaction between two species are independently drawn. In particular the entries of the matrix B had zero mean and there was

no correlation between elements opposed to the diagonal. Allesina and Tang [32] generalized May's condition to different interaction types. In particular they considered the case of a community matrix whose elements are drawn from a distribution with non zero mean and arbitrary correlation between elements opposed to the diagonal. Following May's parametrization, with probability $1-C$ a pair of matrix elements (B_{ij}, B_{ji}) is set equal to zero, while with a probability C the pair is drawn from a given distribution (having each element of the pair the same marginal distribution). In the original paper [32], it was considered the norm of a Gaussian distribution, but the results can be generalized to any distribution. In this case the relevant parameters that shape the distribution of eigenvalues are three: the average of the distribution μ, the variance σ^2 and the correlation between the two elements of the pair τ. May's results reduces to the case $\mu = 0$ and $\tau = 0$.

In the case of $\mu = 0$ the random matrix ensemble reduces to the ones analyzed in section 2.4. One can prove [32] that the effect of a non-zero mean is trivial: it shifts the distribution of eigenvalues and a new eigenvalue (approximately) equal to μCS appears. The other eigenvalues are still uniformly distributed in an ellipse with same sami-axis as in section 2.4, but with a center in $-\mu C$.

Specifically the stability criterion can be expressed in the following way [32, 33]

$$\max\{SE_1, -E_1\sqrt{SE_2(1+E_c)}\} < d\,, \tag{12}$$

where d is the diagonal part of the community matrix (that can be fixed to 1 as done by May), while the other terms correspond to

$$\begin{aligned} E_1 &:= \sum_{i=1}^{S}\sum_{j\neq i}\frac{B_{ij}}{S(S-1)} \\ E_2 &:= \sqrt{\sum_{i=1}^{S}\sum_{j\neq i}\frac{B_{ij}^2}{S(S-1)} - E_1^2} \\ E_c &:= \frac{1}{E_2}\sqrt{\sum_{i=1}^{S}\sum_{j\neq i}\frac{B_{ij}B_{ji}}{S(S-1)} - E_1^2} \end{aligned} \tag{13}$$

.

One can introduce the dependence of E_1, E_2 and E_c on C, μ, σ and τ in these expressions in order to study its direct dependence on those quantities (in particular, in the case $C = 1$, $E_1 = \mu$, $E_2 = \sigma$ and $E_c = \tau$).

Different roles of different interaction types can be obtained directly considering equation 12. Competitive interaction will be characterized by a negative E_1 and (typically) a positive E_c, mutualistic interactions by a positive E_1 and (typically) a positive E_c, while predator-

prey interactions by a close-to-zero mean (which depends on the conversion efficiency, see section 2.7) and a negative E_c. Mutualistic networks are dominated by an eigenvalue $\sim SE_1$, that scales linearly with S and depends only on the average of interactions strengths. For this reason mutualistic interactions are expected to be the less stable ones. Competitive interactions are characterized by a positive correlation E_c which contribute to make them less stable respect to May's case. Predator-prey (and more generally antagonistic ones), being characterized by a very negative correlation [34], are instead typically more stable than what predicted by May's criterion.

The interaction type is determined by its sign and these results are able to take into account this. Another, at least in principle, important aspect that should be taken into account is the structure of the interaction network. In all the case analyzed here, the presence/absence of interactions in the interaction network was randomly and independently drawn. This correspond to choose a Erdös-Rényi structure of the graph describing who interacts with whom. Empirical networks have typically highly non-random features. As explained in section 1.3.1 mutualistic networks are said to be nested, while food-webs have a heterogeneous degree distribution, are hierarchical and interval (see section 1.3.2). All those non-random features, related to network structure, contribute to change the distribution of eigenvalues and then the stability properties of the system. In section 2.6 we review some results obtained studying the effect of empirical (nested) structure on the stability of mutualistic systems. In 2.7 we focus on food-webs, showing new results obtained considering a hierarchical structure of the interaction networks.

2.6 NESTEDNESS AND STABILITY

As explained in section 1.3.1 mutualistic networks do not have a random structure. In many context it has been proposed that nested network were more stable, suggesting a possible advantage of ecological interactions in being nested. In those studies "stability" referred to many different properties of the network structure itself or to an associated dynamics.

When resilience is considered, the results are completely different. The approach is based on the assumption that the community matrix, which governs the dynamics around equilibrium points, have the same structure (in terms of presence-absence of interactions) as the interaction matrix. In particular, nested networks are characterized by larger eigenvalues [11, 18, 32] than their random (ER) counterparts. This negative relation between resilience and nestedness seems very robust.

Resilience defines stability respect to small changes. In chapter 3 we use a global measure of stability, which measures the range of param-

eters that correspond to stable ecosystems. In particular in section 3.8 we study the effect of nested structure. Results depend on parameters, but it emerges quite clearly the secondary role of structure in determining stability.

2.7 STABILITY IN STRUCTURED FOOD-WEBS

Food-webs are known to be very different from random networks. As explained in section 1.3.2 predator-prey networks have a strong directionality, descending biologically from mass constraints in predation. Despite these properties make these webs very different from random networks, in this section we are able to derive a stability criterion for structured food-webs.

In section 1.3.2 we briefly described two models describing the structure of empirical food-webs, the cascade [22] and the niche model [23]. In the former model an interaction network is built by ordering the species and assigning a prey to a predator if the former has a lower rank than the latter (possibly with a given probability). Also in the niche model species are ordered on a proper axis, but the interactions are assigned only if the prey lies in a proper interval around the predator. Niche model is known to perform quite well with data, as it reproduces most of the features observed in empirical networks. In this section we will obtain the distribution of eigenvalues for interaction matrix generated using cascade model. We also show that the eigenvalue obtained in this way well approximates the one of networks generated using the niche model and the empirical ones.

The adjacency matrix K of a network generated via a cascade model can be ordered in such a way that all the elements below the diagonal are null. The cascade model depend only on one parameter determining the number of links present, i.e. the connectance C. We therefore assume that the elements $(A_{ij}, A_{ji})_{i<j}$ of the community matrix A are drawn from a given bivariate distribution if $K_{ij} = 1$ and set to zero otherwise. The random variables (X, Y) drawn from this distribution have given marginal means μ_X and μ_Y, marginal variances σ_X^2 and σ_Y^2 and Pearson correlation ρ_{XY}. As μ_X and μ_Y represent average interaction strengths of a predator-prey system they will have opposite sign, with $\mu_X > 0$ and $\mu_Y < 0$. Moreover we will typically consider the case of $-\mu_Y/\mu_X < 1$, as this ratio is related to the 'efficiency' of predation (measuring the conversion of prey biomass into predator biomass) and the ratio between prey and predator biomass that are both lower than one, with some interesting exception in the ratio between the two biomasses. The goal of this section is to find the dependence of the leading eigenvalue on these parameters.

Let us consider the case $C = 1$ and $\sigma_X = \sigma_Y = 0$. In this case the matrix A is a deterministic matrix having all the elements equal to μ_X above the diagonal and equal to μ_Y below and can be diago-

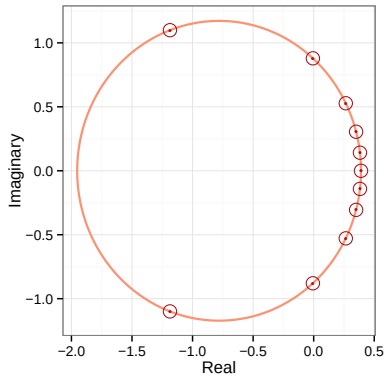


Figure 4: Distribution of eigenvalues of a deterministic matrix generated via a cascade model, with $C = 1$ and $\sigma_X = \sigma_Y = 0$. The eigenvalues lay on a circumpherence in the imaginary plane with center and radius obtained in equations 14 and 15.

nalized exactly. It turns out to have the eigenvalues distributed on a circumference in the imaginary plane centered in

$$c = \frac{\mu_Y - \mu_X(-\mu_Y/\mu_X)^{2/S}}{(-\mu_Y/\mu_X)^{2/S} - 1} \,, \tag{14}$$

and with radius

$$r = \frac{(\mu_X - \mu_Y)(-\mu_Y/\mu_X)^{1/S}}{(-\mu_Y/\mu_X)^{2/S} - 1} \,. \tag{15}$$

Going back to the random matrix we introduced before, we can notice that it can be written as a sum of two matrices $A = D + R$, where D is a deterministic matrix and R a random one. We built D simply by considering a matrix with all the elements on the upper diagonal part equal to μ_U, while all the elements on the lower diagonal part equal to μ_L. We chose μ_U (μ_L) to be equal to the average over the matrix elements of the upper (lower) part the matrix A. In this way the resulting random matrix R will have zero average and given variances on the upper and lower diagonal part.

Since D and R do not commute, the eigenvalues of $A = D + R$ are not a simple combination of the eigenvalues of the two. As we are dealing with random matrices we are not interested in the value of the eigenvalues, but in their distribution. Surprisingly the distribution of eigenvalues of A results as a simple combination of the distributions of eigenvalues of D and R (see fig. 5).

We already know the eigenvalues of D, as they are distributed on a circumference on the complex plane with center and radius obtained in equations 14 and 15, with $\mu_U = \mu_X$ and $\mu_L = \mu_Y$. The distribution eigenvalue of R is not known in fully generality, but it reduces to the elliptic law in the case $\sigma_X = \sigma_Y$. Numerical simulations show that

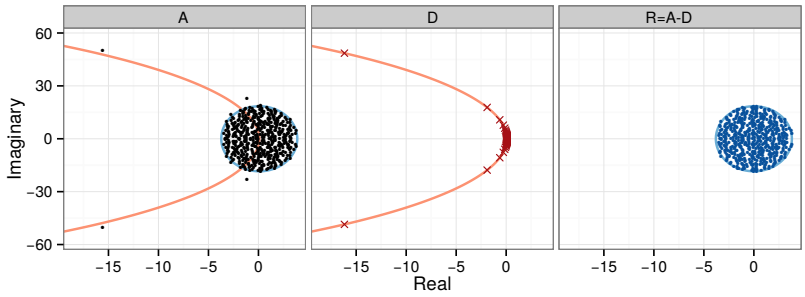


Figure 5: Distribution of eigenvalues of the community matrix A, deterministic matrix D and the random one R, related via $A = D + R$. Despite the fact that D and R do not commute, the distribution of the eigenvalues of their sum seems to be a simple combination of the distribution of the two.

also in the case $\sigma_X \neq \sigma_Y$ the eigenvalues are uniformly distributed in an ellipse. We determined the axis of the ellipse using extensive numerical simulations, finally obtaining the horizontal axis

$$r_h = \frac{\alpha + (S-1)\sigma_L \sigma_U \rho_{UL}}{\sqrt{\alpha}} , \tag{16}$$

and vertical axis

$$r_v = \frac{\alpha - (S-1)\sigma_L \sigma_U \rho_{UL}}{\sqrt{\alpha}} . \tag{17}$$

The value of α, determined using numerical simulations, resulted to be

$$\alpha = \frac{S-1}{12}(\sigma_U + \sigma_L)(\sigma_U + \sigma_L + 4\sqrt{\sigma_U + \sigma_L}) . \tag{18}$$

Using all these results together we are able to obtain the spectral distribution of a matrix generated via a cascade model. Figure 6 show that the distribution and the leading eigenvalue is well predicted by our formulas. One can argue how well the eigenvalue predicted with the cascade model works for other structures, e.g. for the niche model or for the structure of empirical networks. In these cases not all the elements above and below the diagonal are positive and negative respectively. We need therefore to determine the parameters μ_U, μ_L, σ_U, σ_L and ρ_{UL} that best reproduces the matrix using a cascade model. In order to do so, we looked for the order of species having the minimum amount of wrong signs in the matrix and determine the parameters with this ordering. The distribution of eigenvalues and the leading one determined using our approximation performed well also for structures generated using the niche model and for empirical structures (see Figure 6).

Finding the stability criterion for an empirical food-web could seem an impossible goal, as they are far from being random networks. In

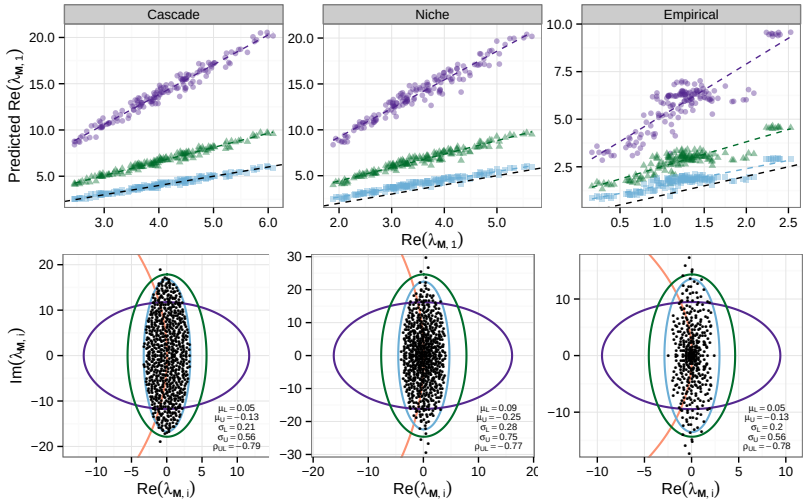


Figure 6: The three panels on the top compare the expected leading eigenvalue calculated using our approximation and the actual one. The three panels on the bottom show the eigenvalue distributions for a particular parametrization. Colors of symbols and lines represent different approximations for the eigenvalues: circular law (purple), elliptical law (green) and the approximation presented in this section. The two panel on the left refer to the cascade model, and show that our approximation works reproducing the eigenvalues. The two on the center compare our prediction with networks generated using a niche model. In the right panel we used the structure of 15 empirical food-webs, each parametrized 10 times. Despite a difference is still present, the stability criterion obtained with our approximation is able to reproduce the eigenvalues obtained parameterizing empirical structures.

this section we have obtained a stability criterion for networks generated via a cascade model, showing that it also well approximates the one valid for empirical webs. To obtain this result we make some approximations one the eigenvalue distribution. This also opens the possibility to deep investigate our hypothesis in the context of random matrix theory.

2.8 DIFFERENT INTERACTION TYPES

The results presented here typically are formulated modeling an ecosystem as a single interaction type was present. Clearly, multiple interaction types shape an ecosystem and their diversity and interdependence could be crucial for stability. To be precise, the original formulation by May (see section 2.1) allows multiple interaction types to coexists in the same ecosystem. In this case, there is no structure within the interaction matrix, and different interaction types are randomly assigned.

It has been proposed [35] that multiple interaction types increase the stability of an ecosystem. In particular it was shown that an ecosystem characterized by two interaction types (e.g. mutualism and predation) was expected to be more stable than an ecosystem with only one interaction type. Moreover it was shown that, with the parametrization used in that work, complexity and diversity increased stability of those communities.

These results seem surprising given the generality of random matrix approach. We investigated [33] the keys assumptions responsible for the results obtained in [35]. In particular the inverse relationship between complexity and stability was a consequence of the parametrization used and not of the presence of multiple interaction types, since the interactions where rescaled such that their average and variance scaled with S and C. This produced the inverse scaling.

On the other side, the observed increase of stability in mixed communities was a consequence of the parametrization used in that work. In particular the authors rescaled the interactions in order to take into account a finite ability of individuals to get resources. The rescaling could be parametrized in many different ways. The observed increase was consequence of this choice, that induced a dependence of the interactions on the intensity of the "mixing" that produced more stable mixed communities. Other choices do not produce this effect.

Also in the case of mixing interaction types, the results obtained with random matrix theory are qualitatively valid. Further results coming from random matrix theory, could make possible to consider more realistic structure and make more precise predictions.

3 STRUCTURAL STABILITY

The World State's motto,
COMMUNITY, IDENTITY, STABILITY.

— Aldous Huxley

3.1 FROM LOCAL STABILITY TO STRUCTURAL STABILITY

In the previous chapter we considered mainly the problem of local stability of ecological systems. Local stability is defined in terms of perturbations on the populations of interest. Assuming that a system is in a stationary state, the question was the response to small perturbation of the stationary state. For small perturbations it is possible to linearize the dynamics around the stationary point and the problem of local stability reduces to find the leading eigenvalue of a Jacobian evaluated at the stationary point.

In this chapter we consider the response of the system to a different type of perturbation. Specifically we consider how the stationary point responds to a change of the parameters defining the dynamics. In particular one is interested in quantifying how much is possible to vary them without changing the dynamical properties of the system. This property of being stable respect to change of parameters is known as structural stability.

In the case of LV equations introduced in section 1.4, one is interested in knowing how much the intrinsic growth rates of species can be varied without varying the property of the system, and in particular without leading any species to extinction. This is in principle a very difficult task, as one has to know the dynamical properties for any value of parameters, in systems where the dimensionality of this parameter space is very large.

The main result of this chapter is a method to calculate the structural stability and the size of the domain of intrinsic growth rates leading to positive abundances. This method allows to quantify the role of interactions strength and structure of interaction network in determining the structural stability. We finally show a new relation between local stability and the response of the system to the change of a single species' growth rate.

3.2 DYNAMICS AND INTERACTIONS OF MUTUALISTIC SYSTEMS

We consider a bipartite mutualistic system described by the following dynamical model

$$\begin{cases} \frac{dN_i^A}{dt} &= N_i^A\left(\alpha_i^A - \sum_{j=1}^{A}\beta_{ij}^A N_j^A + \frac{\sum_{j=1}^{P}\gamma_{ij}^A N_j^P}{1+h_i^A\sum_{j=1}^{P}\gamma_{ij}^A N_j^P}\right) \\ \frac{dN_i^A}{dt} &= N_i^P\left(\alpha_i^P - \sum_{j=1}^{P}\beta_{ij}^P N_j^P + \frac{\sum_{j=1}^{A}\gamma_{ij}^P N_j^A}{1+h_i^P\sum_{j=1}^{A}\gamma_{ij}^P N_j^P}\right) \end{cases}, \quad (19)$$

where N_i^A and N_i^P are the abundances of the animal and plant i, respectively. The other parameters appearing in the equation correspond to the number of animal and plant species (A and P, respectively) the intrinsic growth rates (α_i^A and α_i^B), the intra-guild competition strenghts (β_{ij}^A and β_{ij}^P), the benefits received from mutualistic competiton (γ_{ij}^A and γ_{ij}^P) and the handling times (h_i^A and h_i^P) fixing the saturating constant of the non-linear functional responses. Matrices β^A and β^P and $A \times A$ and $P \times P$ matrices, respectively, while γ^A and γ^P are $A \times P$ and $P \times A$ matrices.

We can rewrite equation 19 in a more compact way, by introducing a vector N, whose components are the abundances of both animals and plants. Specifically, it has $S = A + P$ components, where the first A components are abundances of animals (N_i^A), while the remaining $P = S - A$ are the abundances of plants (N_i^P). We then obtain

$$\frac{dN_i}{dt} = N_i\left(\alpha_i - \sum_{j=1}^{S}\beta_{ij}N_j + \frac{\sum_{j=1}^{S}\gamma_{ij}N_j}{1+h_i\sum_{j=1}^{S}\gamma_{ij}N_j}\right). \quad (20)$$

Both α and h are vector with $S = A + P$ components, where the first A components corresponds to α^A and h^A, while the remaining one correspond to α^P and h^P. The matrices β and γ are both $S \times S$ matrices, with the following form

$$\beta\begin{pmatrix} \beta^A & 0 \\ 0 & \beta^P \end{pmatrix}, \quad (21)$$

and

$$\gamma = \begin{pmatrix} 0 & \gamma^A \\ \gamma^P & 0 \end{pmatrix}. \quad (22)$$

If not explicitly stated, we use the same parametrization used in [37, 38]. All the components of the handling times are set equal to h (i.e. $h_i = h$, $\forall i$). We use a mean-field approximation for the competition parameters, setting $\beta_{ii}^A = \beta_{ii}^P = 1$ and $\beta_{ij}^A = \beta_{ij}^P = \rho$ if $j \neq i$. The mutualistic benefits are parametrized in the following way

$$\begin{aligned} \gamma_{ij}^A &= \gamma_0\frac{a_{ij}}{(k_i^A)^\delta} \\ \gamma_{ij}^P &= \gamma_0\frac{a_{ji}}{(k_i^P)^\delta} \end{aligned}, \quad (23)$$

where a_{ij} is the non-zero block of the adjacency matrix of the interaction network, i.e. $a_{ij} = 1$ if there is an interaction between animal i and plant j, and zero otherwise. k_i^A and k_i^P are the degree (number of interactions) of animal i and plant i respectively, i.e $k_i^A = \sum_{j=1}^{P} a_{ij}$ and $k_i^P = \sum_{j=1}^{A} a_{ji}$. The two remaining quantities, γ_0 and δ, are two parameters quantifying the level of mutualistic strength and the mutualistic trade-off [37].

Therefore, once the interaction matrix (a_{ij}) and the instrinsic growth rates (α) are specified, the model depends on four free parameters: γ_0 (level of mutualistic strenght), ρ (level of interspecific competition), δ (mutualistic trade-off) and h (handling time). In the most part of this section we will consider the case $h = 0$.

3.3 FEASIBILITY AND GLOBAL STABILITY

Let N^* be stationary solution of equation 20, then, by definition

$$N_i^* \left(\alpha_i - \sum_{j=1}^{S} \beta_{ij} N_j^* + \frac{\sum_{j=1}^{S} \gamma_{ij} N_j^*}{1 + h_i \sum_{j=1}^{S} \gamma_{ij} N_j^*} \right) = 0 \qquad \forall i \, . \tag{24}$$

The stationary solution is *feasible* if it corresponds to positive abundances, i.e. $N_i^* > 0 \; \forall i$. Note that, if N_i^* is a feasible solution, then it solves

$$\alpha_i - \sum_{j=1}^{S} \beta_{ij} N_j^* + \frac{\sum_{j=1}^{S} \gamma_{ij} N_j^*}{1 + h_i \sum_{j=1}^{S} \gamma_{ij} N_j^*} = 0 \qquad \forall i \, . \tag{25}$$

A stationary solution is *locally stable* if the system returns to it after any (sufficiently) small perturbation. This condition is equivalent to asking that all the eigenvalues of the Jacobian have negative real-part. A stationary solution is *globally stable* if the system returns to it after any (arbitrarily large) perturbation.

In a Lotka-Volterra (LV) equation with linear functional response response ($h = 0$), we can introduce the matrix $B := \beta - \gamma$ and the Jacobian evaluated at the stationary point (community matrix) has components $J_{ij} = -N_i^* B_{ij}$. If B has all the eigenvalues with positive real part it is said *Lyapunov stable*. The matrix B is *D-stable* if, for any diagonal matrix D with positive entries, the matrix DB is Lyapunov stable. Since in the case $h = 0$ the Jacobian has the form $J = -N^* B$, it follows that *D-stability* implies local stability of any feasible stationary point. The matrix B *Lyapunov diagonally stable* if it exists a diagonal matrix D with positive entries such that all the eigenvalues of the matrix $DB + B^T D$ are positive. It can be proved that Lyapunov diagonal stability implies global stability of *any* feasible equilibrium point.

Lyapunov diagonal stability implies D-stability and D-stability implies Lyapunov stability, while the opposite is not generally true. Moreover, it is generally difficult to verify Liapunov diagonal stability. It

has been conjectured [38] that, using the parametrization introduced in section 3.2, Lyapunov stability implies D-stability and Lyapunov diagonal stability. This statement was proven only in the case $\delta = 0$ or $\rho = 0$, and verified numerically in the most general case. In section 3.3.1 we prove that Lyapunov stability implies D-stability and Lyapunov diagonal stability in the most general case of $\delta > 0$ and $\rho > 0$.

3.3.1 *Proof of conjectures for the global stability of feasible equilibria*

In this section we proof that, using the parametrization introduced in section 3.2, Lyapunov stability implies D-stability and Lyapunov diagonal stability, as conjectured in [38].

Lyapunov stability implies D-stability. This is true not only using parametrization of section 3.2, but for any matrix B representing a system with only competitive or mutualistic interactions. In fact, if a matrix is sign-symmetric (i.e. A_{ij} and A_{ji} have the same sign) then Lyapunov stability implies D-stability [39, 40].

Lyapunov stability implies Lyapunov diagonal stability. One can easily verify that, when parametrized as done in section 3.2, the matrix B is normal, i.e. $BB^T = B^TB$. When a matrix is normal then all its eigenvalues are real and it and its transpose are diagonalized by the same eigenvectors. This implies that B and $(B + B^T)/2$ have the same eigenvalues. If this is the case, then Lyapunov stability implies Lyapunov diagonal stability since we are able to find at least one diagonal matrix D such that $DB + B^TD$ are all positive, this matrix being simply (proportional to) the identity matrix.

The parametrization used here is not the most general one, and one could wonder how much the results obtained in this case are general. To explore the robustness of these results, we considered different parametrization. In particular we drawn the off-diagonal elements β^A_{ij} and β^P_{ij} from a given probability distribution (see below) and parametrized $\gamma^A_{ij} = g^A_{ij}a_{ij}/(k^A_i)^\delta$ and $\gamma^B_{ij} = g^P_{ij}a_{ji}/(k^P_i)^\delta$, where g^A_{ij} and g^P_{ij} are random values and checked weather Lyapunov stability implied Lyapunov diagonal stability.

As B is generally not symmetric (and does not commute with its transpose B^T), the eigenvalues of B and $B + B^T$ are not trivially related. The parameters of the probability distribution defined above, were drawn in such a way that B was Lyapunov stable but $B + B^T$ had at least one eigenvalue with negative real part (otherwise the B would be automatically Lyapunov diagonally stable and then diagonally stable). This last condition corresponds to consider a reactive system [41]. Note that this choice already reduces a lot the space of parameters leading to a non equivalence between Lyapunov stability and Lyapunov diagonal stability. We then tried to find a diagonal

positive matrix D such that $DB + B^T D$ had all the eigenvalues with positive real part, by randomly drawn the diagonal elements D from a specified probability distribution. If such a diagonal matrix was found, then the matrix was Lyapunov diagonally stable. We were always able to find such a matrix with few (< 1000) random steps. We stress that in this case B is very far from being a normal matrix, and our previous demonstration does not hold. This results implies that the equivalence of Lyapunov stability and Lyapunov diagonal stability is valid in a more general scenario for mutualistic systems, that would be interesting to explore in a more systematic way.

To perform this analysis we used the following parametrization. Each pair (β_{ij}, β_{ji}) was set equal to zero with a probability $1 - C_c$ while with probability C_c it was randomly drawn from a pair of lognormally distributed random variables with given mean, variance and correlation. Each pair (g^A_{ij}, g^P_{ji}) was set equal to zero with a probability $1 - C_c$ while with probability C_c it was randomly drawn from a pair of lognormally distributed with given mean, variance and correlation. We tuned the parameters in order to have a matrix $B = \beta - \gamma$ with only positive eigenvalues, and $B + B^T$ having at least one negative eigenvalue. The S diagonal elements of D were drawn from a lognormal distribution with mean equal to one 1 and fixed variance (we used a variance equal to 3). We checked whether $DB + B^T D$ had only positive eigenvalues. If that was the case, then B was Lyapunov diagonally stable, otherwise we considered a new diagonal matrix D and check again the condition. We were always able to find a D such that $DB + B^T D$ had only positive eigenvalues after less than 1000 independent random extractions of D.

3.4 QUANTIFICATION OF STRUCTURAL STABILITY

A system is said structurally stable if a (small) change of parameters does not change its dynamical behavior (i.e. the existence and the stability properties of fixed point, limit cycles and chaotic phases). In our context, the interesting dynamical behavior is a feasible stable fixed point. For the equations introduced in section 3.2 we have proved that feasibility implies stability, and therefore we have to deal only with feasibility.

A possible quantification of structural stability is the volume of parameters leading to feasible stable fixed points. In particular in [38] structural stability has been introduced in the context of Lotka-Volterra equations as the volume of the domain of intrinsic growth rate leading to a feasible and stable equilibrium point.

It is important to observe that if a vector of intrinsic growth rates $\underline{\alpha}$ correspond to a feasible stationary solution, also $c\underline{\alpha}$, being c a positive

constant corresponds to a feasible solution. In a LV equation with linear functional response, the stationary solution is given by

$$N_i^* = \sum_{j=1}^{S} B_{ij}^{-1} \alpha_j \ , \tag{26}$$

therefore the stationary solution corresponding to $c\alpha_i$ is simply cN_i^*. Since c is a positive constant, then also cN_i^* is feasible. This fact implies that the feasibility domain is convex.

Since the feasibility domain is convex, we can quantify it as the solid angle identified by intrinsic growth rate corresponding to positive abundances. This corresponds to quantify the structural stability as the fraction of growth rates vectors that correspond to a feasible solution. We can consider, without loss of generality, only the growth rates lying on the unit sphere. We quantify structural stability as

$$\Xi = \frac{2^S}{Sp_{S-1}} \int d^S\alpha \ 2\|\alpha\|\delta(\|\alpha\|^2 - 1) \prod_{i=1}^{S} \Theta(N_i^*(\alpha)) \ . \tag{27}$$

Up to the factor 2^S this expression exactly quantifies fraction of growth rates vectors on the unit sphere that correspond to a feasible solution.

In this expression we integrate a function on the unit sphere in S dimensions. The integral of a function $f(x)$ on the unit sphere is defined by

$$\int d^S x \ 2\|x\|\delta(\|x\|^2 - 1) \ f(x) \ , \tag{28}$$

The factor Sp_{S-1} is the solid angle in S dimensions, i.e. the surface of a unitary sphere in S dimensions

$$Sp_{S-1} := \int d^S x \ 2\|x\|\delta(\|x\|^2 - 1) = \frac{2\pi^{S/2}}{\Gamma(S/2)} \ , \tag{29}$$

where $\Gamma(\cdot)$ is the Gamma function. The term $2\|x\|\delta(\|x\|^2 - 1)$ that appears in the integrations express the constraint for x of lying on the unit sphere. In equation 27 two other terms appear: 2^S and $\prod_{i=1}^{S} \Theta(N_i^*(\alpha))$. The former is just an arbitrary constant (this choice will be justified later), while the latter expresses the constraint of N_i^* of being positive. The function $\Theta(\cdot)$ is the Heaviside function and it is equal to 1 if the argument is positive and zero otherwise. Therefore $\prod_{i=1}^{S} \Theta(N_i^*(\alpha))$ is equal to 1 if and only if the stationary solution $N^*(\alpha)$ is feasible. $N^*(\alpha)$ is a function of α via equation 25, that in the case $h = 0$ reduces to equation 26. Therefore the integral of $\prod_{i=1}^{S} \Theta(N_i^*(\alpha))$ on the unit sphere quantifies the volume of the domain of intrinsic growth rates corresponding to a feasible solution.

The factor 2^S that appears in equation 27 is an arbitrary choice, and it has been introduced in order to have $\Xi = 1$ when species are

not interacting ($\rho = \gamma_0 = 0$). In the case of $\rho = \gamma_0 = 0$, equation 20 reduces to S uncoupled logistic equations and the stationary solution is feasible and stable if and only if $\alpha_i > 0$.

Unfortunately equation 27 cannot be compute exactly and is very difficult to be evaluated numerically even in the case of linear functional response. In fact to evaluate this integral numerically, we would have to drawn at random intrinsic growth rates on the unit sphere and count the fraction of growth rates corresponding to feasible abundances. As suggested in [38], when S is sufficiently large this evaluation is practically impossible. Fortunatly, equation 27 can be evaluated also in an alternative way, explained in section 3.5.

3.5 FAST AND RELIABLE NUMERICAL CALCULATION OF STRUCTURAL STABILITY

Equation 27 defines Ξ the volume of the domain of intrinsic growth rates leading to feasible solutions. Using the results proved in section 3.3.1 we know that if the matrix B is Lyapunov stable then any feasible stationary point is globally stable. Therefore Ξ is the volume of the domain of intrinsic growth rates leading to feasible and (globally) stable solutions.

Unfortunately the numerical computation of Ξ is unfeasible when the number of species S is sufficiently large. To evaluate the integral in equation 27, e.g. via Montecarlo integration, one should draw at random intrinsic growth rates and count how many of them, out of the total, lead to a feasible equilibrium. In order to have a reliable estimate of this fraction, one should sample the space in such a way that the number of feasible intrinsic growth is much larger than one. This goal would require an exponentially increasing sampling effort as S grows. In this section we provide an alternative, much faster and reliable, way to estimate Ξ.

If $h = 0$, the stationary solution and the intrinsic growth rates are linearly related via $\alpha_i = \sum_{i=1}^{S} B_{ij} N_j^*$. The strategy is to perform a change of variables in equation 27 and integrating over N^* instead of α. Using $\alpha_i = \sum_{i=1}^{S} B_{ij} N_j^*$ we perform the change of variable, obtaining

$$\Xi = \frac{2^S \det(B)}{Sp_{S-1}} \int d^S N^* \; 2\delta\Big(\sum_{ijk} N_i^* B_{ik}^T B_{kj} N_j^* - 1\Big) \prod_{i=1}^{S} \Theta(N_i^*) \,, \quad (30)$$

where $\det(B)$ is the determinant of the Jacobian of the change of variables. The advantage of this passage is that the integration is now on the feasible stationary points and the condition of feasibility is now automatically implemented. Note that it is always possible to perform this change of variable as B is not singular (since it is Lyapunov stable).

The previous expression is still difficult to be evaluated numerically, because of the constraint that appear in the delta function. One can further simplify it by introducing polar coordinates. In particular we write the vector N as $N_i = ru_i$, where r is the norm of N, while u_i are vectors of unit norm. We can perform a new change of variables, passing from N_i to r and u_i. Specifically, for any function $f(N)$ we can write

$$\int d^S\underline{N}\, f(\underline{N}) = \int dr\, r^{S-1} \int d^S\underline{u}\, 2\delta(\|\underline{u}\|^2 - 1) f(r, \underline{u}) \,. \tag{31}$$

Introducing this expression in equation 30 we obtain

$$\Xi = \frac{2^S \det(B)}{Sp_{S-1}} \int_0^\infty dr\, r^{S-1} \int d^S\underline{u}\, 2\delta(\|\underline{u}\|^2 - 1)\, 2\delta(r^2 \sum_{ij} u_i A_{ij} u_j - 1) \prod_{i=1}^S \Theta(u_i) \,, \tag{32}$$

where we used the fact that $\Theta(N_i) = \Theta(u_i)$, since r is positive by definition, and we have introduce the matrix A, defined by $A_{ij} = \sum_k B^T_{ik} B_{kj} = \sum_k B_{ki} B_{kj}$. We can now perform the integration over r obtaining

$$\begin{aligned} &\int_0^\infty dr\, r^{S-1}\, 2\, \delta(r^2 \sum_{ij} u_i A_{ij} u_j - 1) \;= \\ &\int_0^\infty dr\, r^{S-1}\, 2\, \delta\Big(r - \frac{1}{\sqrt{\sum_{ij} u_i A_{ij} u_j}}\Big)\, \frac{1}{2r \sum_{ij} u_i A_{ij} u_j} \\ &= \Big(\sum_{ij} u_i A_{ij} u_j\Big)^{-S/2} \,, \end{aligned} \tag{33}$$

and therefore the integral of equation 27 finally reads

$$\begin{aligned} \Xi = \frac{2^S \det(B)}{Sp_{S-1}} \int d^S u \prod_{i=1}^S \Theta(u_i)\, 2\delta(\|u\|^2 - 1) \Big(\sum_{ij} u_i A_{ij} u_j\Big)^{-S/2} = \\ \frac{2^S \det(B)}{Sp_{S-1}} \int_{S_>} d^S u \Big(\sum_{ij} u_i A_{ij} u_j\Big)^{-S/2} \,, \end{aligned} \tag{34}$$

where $S_>$ means that the integration is carried over the vectors lying on the unit sphere with all the component positve.

As written above, when the integral is written in the form of equation 27 it is unfeasible to evaluate it numerically via Montecarlo integration, since it would require an exponentially increasing sampling to get a given precision. Fortunately, this is not more the case when the integral is written as in equation 34. The main difference is that, after changing variables, one is directly sampling the space of feasible

solutions, without losing computational time in randomly exploring the space of intrinsic growth rates looking for feasible solutions.

In particular Montecarlo integration consists in writing the integral as an average (or a sum) over random points

$$\frac{1}{T}\sum_{a=1}^{T}\left(\sum_{ij} u_i^a A_{ij} u_j^a\right)^{-S/2} \to \frac{1}{Sp_{S-1}}\int_{S_>} d^S u \left(\sum_{ij} u_i A_{ij} u_j\right)^{-S/2}, \tag{35}$$

where $T \to \infty$. In this expression u^a are independent random vectors uniformly distributed on the hypersphere of radius 1 and with only positive components. These two conditions are introduced to satisfy the constraints $\prod_{i=1}^{S}\Theta(u_i)$ and $2\delta(\|u\|^2-1)$ that appear in the integral. T is the sample size, and the average on the left side of equation 35 converges to the right side in the limit of large T.

Clearly one has always a finite sample T used to approximate the integral. It is therefore important to have an estimate of the error made due to a finite sample. Since the left side of equation 35 is an average of a functions over random vectors, this error can be simply estimated using the variance over the sampling, in particular the error is defined as

$$\sigma_{MC} = \frac{1}{\sqrt{T}}\sqrt{\frac{1}{T}\sum_{a=1}^{T}\left(\sum_{ij} u_i^a A_{ij} u_j^a\right)^{-S} - \left[\frac{1}{T}\sum_{a=1}^{T}\left(\sum_{ij} u_i^a A_{ij} u_j^a\right)^{-S/2}\right]^2}. \tag{36}$$

3.6 POSSIBLE BIASES IN PREVIOUS ANALYSIS OF STRUCTURAL STABILITY

In section 3.5 we showed how it is possible to estimate numerically the structural stability in a fast and reliable way. This method allows to compute the size of the domain of conditions leading to a stable and feasible equilibrium. In previous approaches [38], structural stability was not directly calculated, but approximately inferred using a regression method. Using the first approach we obtain partially different results respect to the ones obtained in [38]. In this section we show why the method used in [38] could be biased and not always applicable.

The method proposed in [38] was based on structural vectors. The authors defined it as the center of feasibility domain and it was calculated by transforming the mutualistic dynamics in an effective competitive dynamics. Using this effective dynamics it was possible to calculate an effective structural vector, that was then transformed back to the one of the mutualistic system. Starting from the structural vectors the authors considered different perturbations (by generating different growth rate vectors at a given angular difference with the

structural vectors). The dynamics was then integrated and the probability that all the species survived given an angular perturbation was calculated. Running this several times and for several different perturbations, it was possible to perform a regression between the probability, the angular perturbation, nestedness and other parameters. Using the coefficient calculated with the regression, one was then able to calculate the effect of nestedness and other parameters on structural stability.

Here we present some possible issues emerging using that approach. We do not investigate explicitely which one of them determines the differences we observe. Some of these issues could be relevant, while others could be irrelevant for the practical goals they were introduced for.

3.6.0.1 *It is not always possible to find the structural vector*

In order to calculate the structural vector, one needs to transform the mutualistic system into an effective competitive one. In order to do so one defines the matrix $T = 1 + \gamma\beta^{-1}$, and multiplies both side of the equation

$$\alpha_i = \sum_{j=1}^{S} B_{ij} N_j^* \,, \tag{37}$$

by T. The effective interaction matrix is then $B^{eff} = TB$. In order to calculate the structural vector one has to assume that the eigenvector associated with the largest eigenvalue has only positive components. This is not generally true. Since this assumption does not hold in general, it is not always possible to calculate the structural vector. In section 3.9.1 we propose an alternative method. Notice that our approach (see section 3.5) does not need to calculate the structural vector.

3.6.0.2 *The structural vector is not the center of the feasibility domain*

When is possible to calculate the structural vector for the effective competitive system, one has to transform it back to a vector of the mutualistic system. The structural vector was the center in the feasibility domain of the effective system. For volume in more than two dimension a "center" could be defined in different ways, if the volume has not particular symmetries. The transformation from the effective to the original system is not a rotation (since it is just a multiplication the matrix by T^{-1}) and therefore it does not preserve the angles between vectors. If a vector is the center of the feasibility domain in the effective system, it will be not in general the center in the original domain. In particular its distance respect to the actual center of original domain will be dependent on parametrization and network structure, as the transformation matrix depends on them.

3.6.0.3 *The regression procedure can in principle produce biases*

The relation between structural stability and network structure was obtained by calculating the probability of coexistence given an angular perturbation respect to the structural vector and was then performed a regression between the logit of the probability and the angle of perturbation, the parameters ρ, γ, δ and nestedness. In order to perform a regression, one has to assume a functional dependence. The functional dependence used to perform the regression cannot be justified a priori. An uncorrect functional dependence produce a dependence of the fitted parameters on the parameters, via different data that are used to perform the fit and therefore it cannot be reliable in the poorly sampled regions. Moreover it cannot provide a method to calculate the structural stability for a given network and parametrization, as one can calculate only the probability of coexistence given an angle of perturbation.

3.7 DEPENDENCE OF STRUCTURAL STABILITY ON STRUCTURE, INTERACTION STRENGTH AND ASYMMETRY

The volume of the feasibility domain depends on all the parameter that we consider in the definition of the model. In this section we analyze the dependence of Ξ on γ, ρ and δ. Finally we study how it depends on nestedness.

Figure 7 show the dependence of structural stability Ξ on interaction strengths γ and ρ and on mutualistic trade-off δ.

3.8 STRUCTURAL STABILITY OF EMPIRICAL MUTUALISTIC NETWORKS

Using the approach explained in section 3.5 we can directly compute the structural stability Ξ of any given network. In particular we can study the effect of structure of empirical mutualistic networks on structural stability.

Empirical networks are different from random networks 1.3.1. In particular they are more nested than random counterparts. In section 3.7 we explore directly the effect of nestedness on structural stability. In this section we use empirical network, and compare their structural stability with respect to the structural stability of their randomizations.

We randomize empirical networks in two ways. We randomize only the structure of the underlying networks, and the corresponding interaction matrices are obtained using the parametrization explained in section 3.2.

ER randomization. In this case we conserve the numbers of species (S_1 and S_2) and the number of links (L), generating Erdős-Rényi ran-

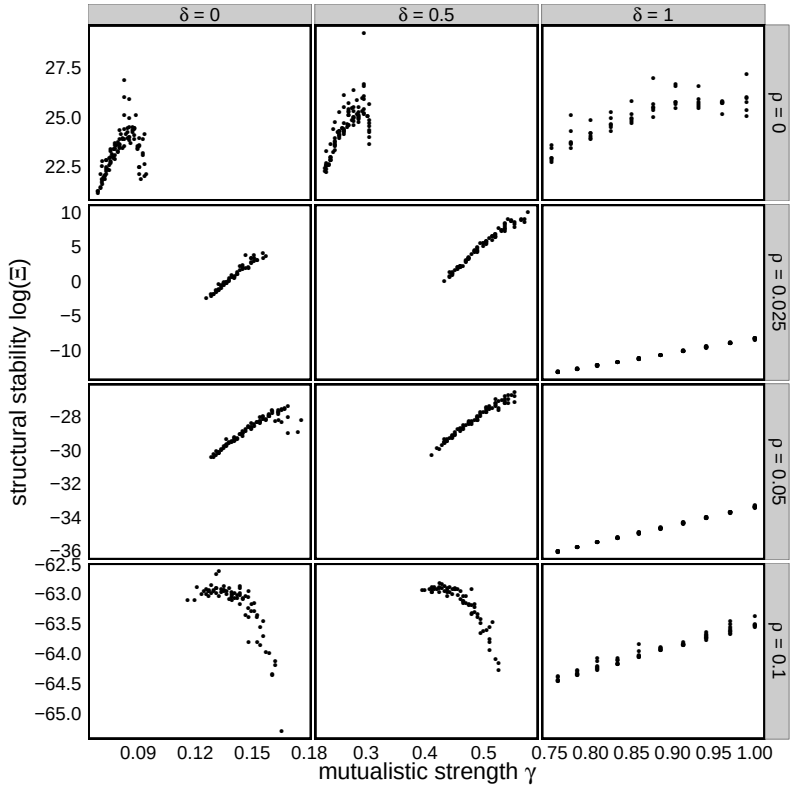


Figure 7: Structural stability Ξ of random networks vs. mutualistic strenght γ and different values of δ and γ.

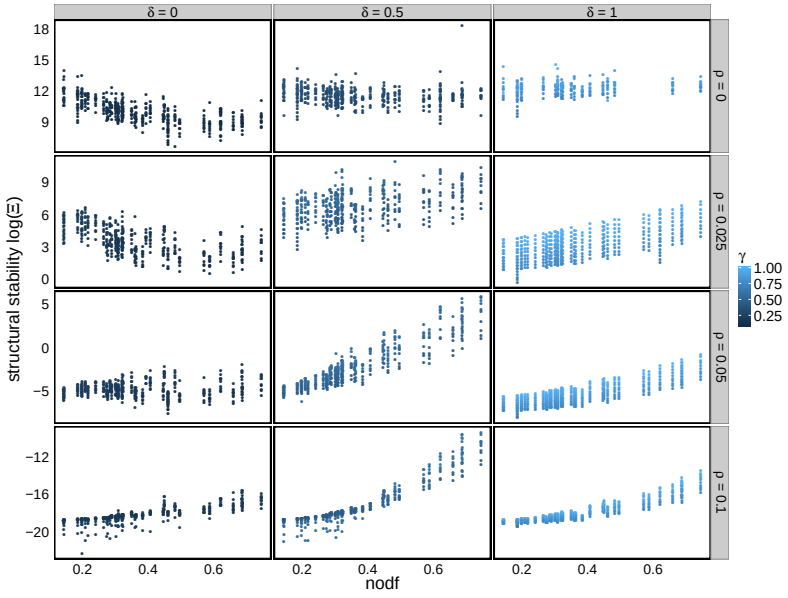


Figure 8: Structural stability Ξ of random networks vs. nestedness (nodf) and different values of δ and γ.

dom graphs. Among all the possible graphs we keep only the graphs that are fully connected (i.e. they have only one connected component). The graphs generated in this way are connected graphs having the same S_1, S_2 and L of the corresponding empirical networks.
FD randomization. In this case we conserve the numbers of species (S_1 and S_2) and the degree on all the nodes in the network (i.e. we are conserving the exact degree sequences of the bipartite graph) and we require the graph to be fully connected. In order to generate these randomized networks we used the Curveball algorithm [42] adding a condition on the connectedness of the graph. Specifically we check connectedness at every move, accepting it only if it does not disconnect the graph. The relaxation time (i.e. the number of randomizations moves) was chosen to be $10\min(S_1, S_2)$ (in [42] was shown that $5\min(S_1, S_2)$ moves were enough).

In order to determine the system, we have to specify three parameters: the competition strength (ρ), the mutualistic strength (γ), and the mutualistic trade-off (γ). As shown in section 3.7.

3.8.1 *Parametrization of empirical mutualistic networks*

We ran 100 simulations for each empirical network and each triplet of values (ρ, δ and γ_0). We considered two values for each parameter ($\rho = 0, 0.4$ and $\delta = 0, 0.5$), while the mutualistic strength was set to values such that the interaction matrix is Lyapunov stable. As explained in [38], as γ_0 increases the real part of the leading eigenvalue decreases and eventually crosses zero, resulting in an unstable matrix. Since this critical value of the mutualistic strength γ_0^c depends on the network structure, we chose two different values of the mutualistic strength, accordingly to the procedure explained in the following. We start from a value of the mutualistic strenght large enough to correspond to an interaction matrix that is not Lyapunov-stable. Then we decreased γ_0 by multiplying it to a constant factor lower than one (in the simulations shown here we used 0.9). We stop when the value of γ_0 found had crossed γ_0^c leading to a stable matrix. This value was therefore guaranteed to lie between $0.9\gamma_0^c$ and γ_0^c. We ran simulation for the value of γ_0 found in this way and for a value equal to it divided by 10. In these way all the simulation were ran for two values of γ_0: one close and one much smaller than the critical value γ_0^c. We call these two values γ_a and γ_b, where γ_a is guaranteed to lie between $0.9\gamma_0^c$ and γ_0^c, while γ_b belongs to the interval $[0.09\gamma_0^c, 0.1\gamma_0^c]$.

It is very important to notice that γ_a and γ_b were obtained using the empirical networks. Randomized version of them have in principle very different values of γ_0^c. Even if we expect random networks to be more stable than the empirical ones [11, 18], it was possible that a randomized version of the original network was not stable using γ_a and γ_b. Since the structural stability has no meaning for an unstable

system, we keep track of those network but we did not compute the value of Ξ and these networks are not shown when the empirical data are compare with randomized networks.

3.8.2 *Results*

For each parametrization (ρ, δ and γ_0), empirical network and randomization procedure (ER or FD) we ran 1000 randomizations. We compute for each matrix (empirical or randomized) the value of Ξ and the corresponding error σ_Ξ as explained in section 3.5. Since the distribution of Ξ across randomizations could be skewed, as done in section 3.7, we considered $\log_{10}\Xi$ (and the corresponding error) and we compared it with the value of $\log_{10}\Xi_e$ obtained with the empirical networks. Two compare the values obtained with the randomization and the empirical one, we compute both the $z-$score and the p value.

The $z-$score was computed by calculating the average value of structural stability over randomizations: $\langle\log_{10}\Xi\rangle$ and its corresponding variance $\langle(\log_{10}\Xi)^2\rangle - \langle\log_{10}\Xi\rangle^2$

$$z = \frac{\log_{10}\Xi_e - \langle\log_{10}\Xi\rangle}{\sqrt{\langle(\log_{10}\Xi)^2\rangle - \langle\log_{10}\Xi\rangle^2}} \,. \tag{38}$$

We computed the p-value as the probability to find in the randomization a value larger (smaller) than the empirical one. The empirical value was consider significantly larger (smaller) than the ones obtained with the randomizations if it was in the larger (smaller) than 97.5% of data. This correspond to choose a two-tailed threshold for the p value equal to 0.05.

Note that the value of the $z-$score depends on the choice of using $\log_{10}\Xi$ instead of Ξ, while the p-value, is irrespective of this choice.

Figure 9 and 10 show the results obtained using ER and FD randomization respectively. In these two figures we reported only the results for which $\sigma_\Xi < 0.1\Xi$, to select for networks with very reliable results.

When compared with ER randomizations (see Figure 9), the empirical networks can be more or less structurally stable depending on parameters. In particular only in the cases $\rho = 0$, $\delta = 0$ and $\gamma = \gamma_b$ empirical networks are significantly more structurally stable than ER random networks. Moreover there are other cases where empirical networks are *less* structurally stable than random counterparts.

FD randomizations (see Figure 10) show a much clearer scenario. In all the cases, with very few exceptions, empirical networks are not more structurally stable than random counterparts. This fact implies that any difference between empirical and random networks is a consequence of the degree sequence and not of any other structural difference (such as nestedness).

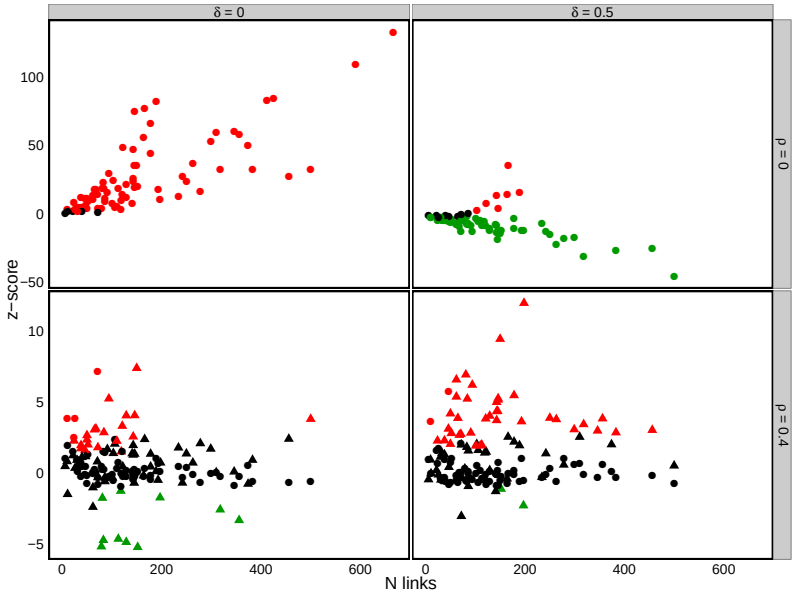


Figure 9: Structural stability Ξ of empirical networks and ER randomizations. $z-$score is computed as defined in equation 38 and is plotted vs. the number of links of the network. Each panel represents a combination of values of δ and ρ. Each symbol is a network, circles represents $\gamma_0 = \gamma_b$ while triangles correspond to $\gamma_0 = \gamma_a$. Colors represent significance of difference with respect to randomizations: red stands for Ξ significantly larger than randomizations, green means significantly smaller while black symbols are not significant. Significance was computed with a two-tailed threshold equal to 0.5. In this figure are reported only empirical network for which $\sigma_\Xi < 0.1\Xi$.

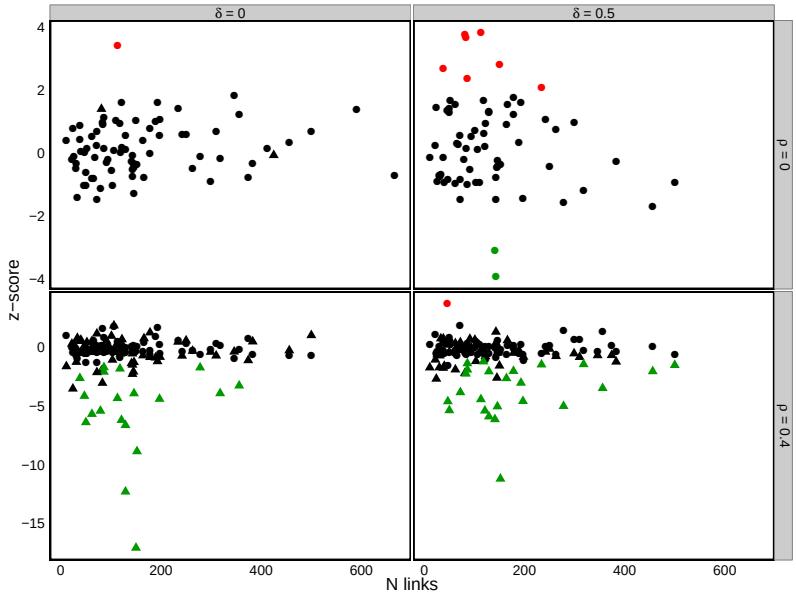


Figure 10: Same as in Figure 9, but using FD randomizations. In this figure are reported only empirical network for which $\sigma_\Xi < 0.1\Xi$.

3.9 DIRECTION OF MINIMAL STRUCTURAL STABILITY: DEFINITION AND CALCULATION

In section 3.4 we show how the volume of the domain of the intrinsic growth rates leading to a feasible equilibrium could be written as an integral. This volume Ξ is not the only interesting quantity that is worth to compute. In particular one could think at how the system respond when the intrinsic growth rate of a single species is modified. The volume Ξ quantifies how large is the total volume of feasible intrinsic growth rate, in this section we introduce the quantity ξ_i, that measures how much the growth rate of the species i can be modified without having extinctions (i.e unfeasible populations).

In [38] the authors introduced the structural vector as a "center" of the domain of feasible intrinsic growth rates. As explained in section 3.6, the calculation of this vector is in general an issue, since it cannot be simply computed from the interaction matrix. The method we introduced in section 3.4 and 3.5 does not require the knowledge of this vector. In order to study ξ_i we need to define and compute the structural vector for any given matrix and parametrization. In [38] the structural vector was defined as the "center" of the domain of feasible intrinsic growth rates. There are several possible way to define a center of a hypervolume and, without specific assumptions, they are all different. The most natural choice is to define the structural vector as the barycenter of the domain of feasible intrinsic growth rates. In section 3.9.1 we explain why this is a convenient and reasonable def-

inition and how to compute the coordinates of the barycenter in the case of study.

For what follows it is necessary to assume that the domain of feasible intrinsic growth rates is a convex volume. This statement is always true in the case of a linear functional response. It is trivially true since any linear combination of feasible intrinsic growth rates results in a feasible intrinsic growth rate (because of linearity of the equation).

Let α^c be the structural vector, defined as in section 3.9.1. The quantity ξ_i is defined as

$$\xi_i = \frac{1}{2\pi}\int_0^{2\pi} d\theta \prod_{j=1}^{S} \Theta(N_j^*(\tilde{\alpha}(\theta, i))) \,, \tag{39}$$

where $\tilde{\alpha}(\theta, i)$ is a vector in the intrinsic growth rate space, whose components are defined as

$$\tilde{\alpha}_j(\theta, i) := \alpha_j^c \cos(\theta) + \delta_{ij} \sin(\theta) \,. \tag{40}$$

The quantity θ is the angle between the structural vector α_j^c and the perturbed growth rate vector. The integral 39 quantifies the maximal pertubation angle θ, or more precisely, the size of the region of θ's corresponding to a feasible equilibrium. In particular ξ is always lower than one. In the case of a linear functional response it is always lower than $1/2$. In fact, in this case, since the equation is linear, if an intrinsic growth rate correspond to a feasible equilibrium its reciprocal does not.

The quantity ξ_i is different for different species. More robust species (respect to changes in growth rates) will correspond to the ones having larger ξ_i, while the less ones are characterized by a small ξ_i. When ξ_i approaches 0, there are not values of intrinsic growth rates that can sustain a positive population of the species i.

3.9.1 *Computation of the structural vector*

We define the structural vector as the barycenter of the domain of feasible intrinsic growth rates. There are several properties that are valid in full generality. Any plane passing from the barycenter divides the volume in two subvolumes of equal size.

The coordinates of the barycenter can be found as the center of mass of the volume considered. In our case, the i-th component of the center of mass would be

$$\alpha_i^c := \frac{\int d\alpha \; 2\delta(\|\alpha\|^2 - 1) \; \prod_{j=1}^{S} \Theta(N_i^*(\alpha)) \; \alpha_i}{\int d\alpha \; 2\delta(\|\alpha\|^2 - 1) \; \prod_{j=1}^{S} \Theta(N_i^*(\alpha))} \,. \tag{41}$$

This integrals are generally difficult to be computed and can be in principle be evaluated in a very similar way to the one we introduced in section 3.5. We instead used a different (and faster) approach.

The basic idea behind this second approach, was to generate a random-walk on the domain of intrinsic growth rates and calculate the barycenter averaging over all the positions visited in the random-walk. These approach coincides with evaluation of the integral in equation 41 since the random-walk covers uniformly the space. To be more precise a random-walk in a confined volume does not occupy uniformly the space. In order to avoid this problem we imposed periodic boundary conditions. Specifically, the random-walk was implemented in the following way. An initial vector $\alpha(0)$ of intrinsic growth rate was considered such that it corresponded to a feasible population. We randomly drawn a direction (a random vector $\alpha_i^r(0)$ orthogonal to the first one). We move the original vector of a small angle in that direction

$$\alpha_i(1) = \alpha_i(0)\cos(\eta) + \alpha_i^r(0)\sin(\eta)\ , \tag{42}$$

where η is a Gaussian random variable with small variance. If the new vector $\alpha_i(1)$ correspond to a feasible equilibrium, then we accept the move. If it does not, we have to implement periodic boundary conditions. Specifically, given a vector $\alpha_i(0)$ that belongs to the feasibility domain and a random vector $\alpha_i^r(0)$, we can calculate the conditions that the perturbation need to satisfy in order to correspond to a feasible equilibrium. Let assume that to be feasible the perturbation η must belong to an interval $[\eta_{min}, \eta_{max}]$, with $\eta_{min} < 0$ and $\eta_{max} > 0$. We stress that these two values, $\eta_{min} < 0$ and $\eta_{max} > 0$, depend on both $\alpha_i(0)$ and $\alpha_i^r(0)$. We impose periodic boundary conditions simply by considering

$$\eta' = \eta - \eta_{min} + [(\eta - \eta_{min}) \bmod (\eta_{max} - \eta_{min})]\ , \tag{43}$$

where mod is the modulo operation. Considering η' we are guaranteed to perform a random walk inside the of feasible intrinsic vectors.

We simulate this dynamics, using equation 42, and we compute the barycenter as average over vectors $\alpha_i(t)$. Since the dynamics of equation 42 produces correlated vectors, we introduce a relaxation time of the random walk τ and we include in the average only vectors with a temporal lag equal or larger than τ. In particular

$$\alpha_i^c = \frac{1}{N}\sum_{m=1}^{N} \alpha_i(m\tau)\ , \tag{44}$$

where N is the number of independent vector considered. In all the results shown in this work we considered $N = 200S$ and $\tau = 5S$ (or larger values), where S is the total number of species.

3.10 DEPENDENCE OF MINIMAL STRUCTURAL STABILITY ON PARAMETRIZATION

In section 3.9 we introduced the structural stability ξ_i of a single species i, that measured the range of growth rates leading to a posi-

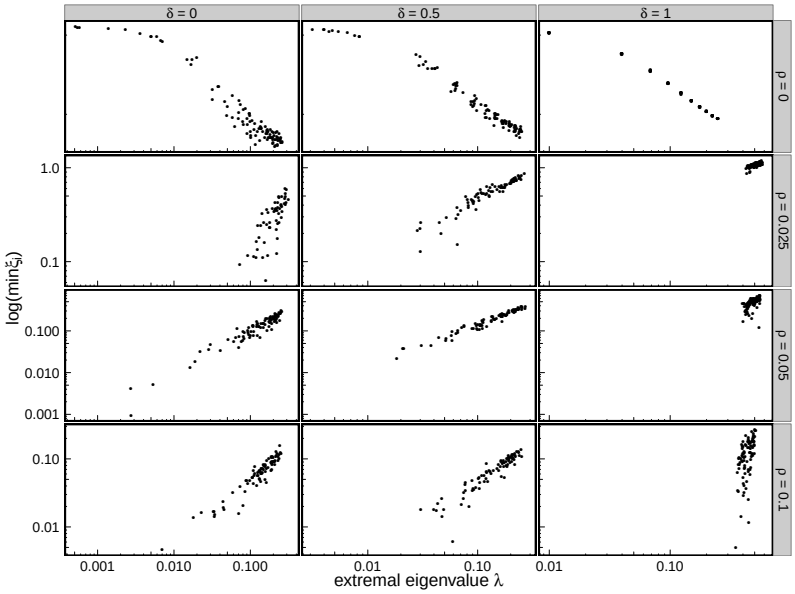


Figure 11: Dependence of minimal structural stability on extremal eigenvalue. The structural stability of a single species ξ_i, defined in section 3.9, quantifies the range of growth rates leading to a feasible population of the species i. Here we plot this value for the species having the smallest value vs the smallest eigenvalue of the matrix B, which determines the Lyapunov stability of the system. Each inset refers to a given combination of the competition ρ and the mutualistic trade-off δ. Different points where obtained varying lambda. The interaction matrix γ where drawn as independent 50×50 random matrices with 500 links. The eigenvalue was computed as the minimal eigenvalue of the corresponding interaction matrix B obtained with the parametrization discussed in section 3.2. Each point is an average over 10 random independent realizations.

tive population of the species i. It is of particular interest the species having the smallest value of ξ_i as it represents the species at most risk or the direction of minimal structural stability.

In principle this minimal value of ξ_i depends on all the parameters in the model. Figure 11 shows that it is striclty related to an a-priori very different quantity. In fact it is very correlated to the value of the smallest eigenvalue of the interaction matrix, which is the quantity determining the Lyapunov stability of the system.

This result shed a new light on the role of the eigenvalue in determining the stability. As the eigenvalue crosses zero, also the minimal structural stability tends to zero.

3.11 CONCLUSION

In this chapter we have studied the structural stability of large mutualistic ecosystems. We have introduced a method to calculate directly this quantity and determined the effect of network structure and interaction strength on it.

The effect of network structure strongly depends on the values of interaction strengths. Empirical network can be more or less structurally stable depending on which between competition and mutualism dominates. Moreover this effect solely depends on the degree sequence of the network and not on further structural properties of empirical networks.

We have then explicitly studied the effect of nestedness on structural stability by generating synthetic networks with arbitrary nestedness showing that their relation is strongly dependent of the interaction strength and the effect of structure is negligible.

Structural stability depends on the response of all the system to the change of parameters. We have then considered the case where only one species was affected. In this case we quantify for each species the range of growth rates leading to a feasible solution. Interestingly we found a strong relation between the minimal range across the species and the eigenvalue of the Jacobian determining the local stability properties of the system. This relation suggests a deep connection between local stability and robustness of the system to perturbations of parameters.

4

PERSISTENCE OF A METAPOPULATION IN A RANDOMLY FRAGMENTED LANDSCAPE

Remember, remember always that all of us,
and you and I especially,
are descended from immigrants and revolutionists.

— Eric Hoffer

4.1 METAPOPULATION THEORY

We live in a fragmented patchy world [44]. The individuals of a species do not occupy space homogeneously, since the spatial locations that support life are discrete and fragmented. The typical example is an archipelago and a population of some terrestrial species living on islands. A similar scenario can be also found in a forest, where the resource distributions in not uniform. Moreover human activity typically increase the fragmentation of the landscape. The term metapopulation refers to a population living in such fragmented landscapes, composed of isolate patches. The main goal of metapopulation theory is to relate the level of fragmentation in the landscape to the persistence (or extinction) of the population considered.

We can imagine to consider a single species, with subpopulations that potentially can be present in each patch. The main assumption of metapopulation is that patches are isolated, i.e. the typical timescale of dispersal between patches is much larger than the timescale involved in the local population dynamics (e.g. the typical lifetime of an individual). Under these assumptions, one can model the dynamics of a metapopulation separating the two timescales. On the shorter timescale, the population in each patch follow its dynamics. We can imagine that, due to fluctuations, the subpopulation in a single patch can go extinct. If there was not dispersal, the subpopulations in single patches go extinct one by one, and in a finite time the whole population would go extinct. Dispersal between patches, by repopulating empty patches, can balance local extinctions, and, if it is strong enough can avoid global extinctions.

The problem of persistence is therefore a problem of balance between local extinctions and dispersal. In subsection 4.1.1 we will introduce a dynamical model for metapopulations [45, 46, 47], that specify in more rigorous term how to quantify this balance. This model, and all its possible generalizations, requires the specification of how the landscape is fragmented (e.g. in the archipelago case, the positions and the sizes of all the islands). The goal of this chapter is to deter-

mine the condition for persistence of a metapopulation in a *random* fragmented landscape, i.e. where positions of patches are randomly distributed in space and their sized drawn from a probability distribution. We show how this problem can be mapped into the problem of finding the expectation of the leading eigenvalue of a random matrix ensemble, known as Euclidean Random Matrices. We are able to derive a persistence criterion showing that a handful of quantities drives persistence: the density of patches, the variability in their value, the shape of the dispersal kernel, and the dimensionality of the landscape. We also demonstrate that metapopulations close to extinction are spatially localized, with few patches clustered in space contributing disproportionately to persistence.

4.1.1 *Metapopulation dynamical models*

The starting point for metapopulation models is Levins' model [48]. The model assumes a large number of identical patches of suitable habitat that may or may not be occupied by a local population. The proportion of patches $p(t)$ that are occupied at time t is governed by the equation

$$\frac{dp(t)}{dt} = cp(t)\,(1-p(t)) - ep(t)\,, \tag{45}$$

where c is the colonization rate and e is the extinction rate. The metapopulation is persistent whenever the stable solution of the equation above is positive. Levins' model has two stationary solutions: $p^* = 0$ and $p^* = 1 - e/c$. The latter is a positive stable solution, i.e. the metapopulation is persistent, if $\delta := e/c < 1$.

This model is spatially implicit: individual patches and local population persistence probabilities are not modeled explicitly. To make the model spatially explicit, one should specify the location of the patches in the landscape. The geometry of the landscape will then dictate the rate at which individuals in a patch colonize other patches, with colonization rates typically decreasing with the distance between the patches. Moreover, patches can differ in size or suitability such that the dynamics of a local population depends on the patch "value".

Hanski & Ovaskainen [45, 46, 47] proposed a model in which the colonization and extinction rates depend both on the particular patch under consideration and the occupancy state of all the other patches:

$$\frac{dp_i(t)}{dt} = (1-p_i(t))\,C_i - p_i(t)E_i\,, \tag{46}$$

where $p_i(t)$ is the probability that patch i is occupied at time t, and $C_i = C_i\,(p_1(t),\ldots,p_N(t))$ and $E_i = E_i\,(p_1(t),\ldots,p_N(t))$ are the colonization and extinction rates of patch i as a function of the current

distribution of the population over the landscape (N is the total number of patches).

The model also incorporates the notion of patch "value" (e.g., a function of the patch size and the density of available resources), which measures the probability of incidence on a given patch without any immigration from other patches, and can therefore be interpreted as the carrying capacity of the patch. The colonization rates are assumed to be directly proportional to this value, while the extinction rates are assumed to be inversely proportional to it. Denoting the value of patch i by A_i, the extinction rate in patch i is $E_i = \delta/A_i$, where δ is the extinction rate per unit patch value. Without loss of generality we assume that, up to a rescaling of δ, the mean patch value (averaged over all the N patches) is one, i.e., $\sum_i A_i/N = 1$.

The colonization rate of patch i is composed of the contribution of all other patches, with the contribution of each patch being a function of the distance between the patches. This function is called *dispersal kernel* and is denoted $f\left(|x_i - x_j|/\xi\right)$, where x_i is the spatial position of patch i, $|x_i - x_j|$ is the Euclidean distance between patches i and j. In the following we will consider the case of f depending on one parameter with length dimensions, specifically we will consider the case $f(|x_i - x_j|)$, where ξ is the *dispersal distance*, i.e., the characteristic distance the population can traverse in a unit of time. The colonization rate is therefore $C_i = \sum_{j\neq i}^{N} f\left(|x_i - x_j|/\xi\right) A_j p_j$. Hanski & Ovaskainen [45] studied the special case of an exponential dispersal kernel $f\left(|x_i - x_j|/\xi\right) = \exp\left(-|x_i - x_j|/\xi\right)$, but for a general kernel, we have:

$$\frac{dp_i(t)}{dt} = \sum_{j\neq i}^{N} (1 - p_i(t))\, f\left(\frac{|x_i - x_j|}{\xi}\right) A_j p_j(t) - \frac{\delta}{A_i} p_i(t)\,. \qquad (47)$$

This equation can be obtained from a stochastic process describing the occupation of the N patch. One can indeed consider N random variables σ_i, taking value 1 if the species of interest is present in patch i and 0 if it is not. The stochastic dynamics is fully specified once transition rates are specified. In this case one can imagine that in each patch the σ goes from 1 to 0 with a rate proportional to δ (and depending on patch value), while it flips to 1 depending on a rate of migration and the occupancy of other patches. This process is known as contact process [49] in stochastic dynamics literature or as SIS model in epidemiology. One can obtain a deterministic equation for $\langle\sigma_i\rangle$, which depends on the correlations $\langle\sigma_i\sigma_j\rangle$. If we neglect the correlations, assuming $\langle\sigma_i\sigma_j\rangle = \langle\sigma_i\rangle\langle\sigma_j\rangle$, we obtain a close set of equations for $\langle\sigma_i\rangle$ that turn out to have exactly the same form of equation 47, with $p_i = \langle\sigma_i\rangle$.

Introducing the matrix M with the definition

$$M_{ij} = \begin{cases} f\left(\frac{|x_i - x_j|}{\xi}\right) A_i A_j, & i \neq j, \\ 0, & i = j, \end{cases} \tag{48}$$

Hanski & Ovaskainen [45] derived a simple persistence rule: the metapopulation persists as long as the extinction rate $\delta < \lambda$, where λ is the leading (largest, rightmost) eigenvalue of M. Thus, the criterion for metapopulation persistence has the same form found in Levins's model, with the leading eigenvalue λ (dubbed the "metapopulation capacity" [45]) playing a key role for persistence.

4.2 DISPERSAL MATRIX AS A RANDOM MATRIX

4.2.1 *Euclidean Random Matrices*

Euclidean Random Matrices [50] were introduced to study the large class of physical system where disorder has a spatial origin. The vibration spectra of glasses [51] or instantaneous normal modes in liquids [52] are just two of the possible applications.

Consider a region Ω in a d-dimensional Euclidean space. Take a set of N points x_i randomly and uniformly distributed over this region (the patches). Let $f(x_i - x_j)$ be an arbitrary function of any two points x_i, x_j in Ω. Then the matrix A with $A_{ij} = f(x_i - x_j) - u\delta_{ij} \sum_k f(x_i - x_k)$ is called a *Euclidean Random Matrix* [50]. The function f is typically considered to depend only on the distance between x_i and x_j: $f(x_i, x_j) = f(|x_i - x_j|)$.

The spectrum of an Euclidean random matrix has not been fully solved. It is possible to obtain two approximation, expanding the solution at low and high density respectively.

In the following we will consider only the case $u = 0$. Sometimes, instead of working with the Euclidean Random Matrix A, we consider the matrix $B = A - f(0)I$, where I is the identity matrix. Since f is assumed to depend only on the distance between points, all diagonal elements of A are equal to $f(0)$. Therefore the only difference between A and B is that the latter has its diagonal entries set to zero. The effect of this is simply to shift the eigenvalues by a constant: if

$$Aw_\alpha = \lambda_\alpha w_\alpha \tag{49}$$

(where w_α, λ_α are the αth eigenvector and eigenvalue of A, respectively), then

$$Bw_\alpha = (A - f(0)I)\, w_\alpha = (\lambda_\alpha - f(0))\, w_\alpha. \tag{50}$$

meaning that the eigenvectors of B are the same as those of A, and that the eigenvalues of A are those of B shifted by $f(0)$.

In the special case when f is equal to one for $|x_i - x_j| < \xi$, and to zero otherwise (i.e. a theta function, where ξ is parameter), the Euclidean Random Matrix generated this kernel is called a *Random Geometric Graph* [53]. We sometimes make use of Random Geometric Graphs, as they provide the most intuitive case of Euclidean Random Matrices, and they can be seen as the adjacency matrix of an undirected graph.

4.2.2 *The dispersal matrix*

The dispersal matrix M (Eq. 48) has special properties as long as the dispersal rate depends only on the distance between patches, which we will assume throughout. In this case M is symmetric ($M_{ij} = M_{ji}$) and nonnegative ($M_{ij} \geqslant 0$). Symmetry implies that all the eigenvalues of M are real. Then, as long as M is irreducible (a condition typically met by the matrices studied here), M will have a unique largest eigenvalue [54, 55], which is the leading eigenvalue λ referred to above. Moreover, the eigenvector associated with this eigenvalue can be chosen with strictly positive components. Note that, since we are interested in the largest eigenvalue overall as opposed to the largest in absolute value, primitivity of M is not required.

Moreover, using the symmetry of M, one can derive a lower bound for the magnitude of λ. Since by definition

$$\lambda = \max_{\theta_i \neq 0} \left(\frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \theta_i M_{ij} \theta_j}{\sum_{k=1}^{N} \theta_k^2} \right), \tag{51}$$

we may arbitrarily set $\theta_i = 1$ for every i to obtain

$$\lambda \geqslant \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} M_{ij} \tag{52}$$

(e.g., [56]). This sum can be interpreted as the arithmetic mean of the row sums of the matrix M. Therefore, the average row sum provides a lower bound for the leading eigenvalue of symmetric matrices.

This approximation is in fact a special case of the more general formula

$$\lambda \geqslant \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} (M^q)_{ij}}{\sum_{i=1}^{N} \sum_{j=1}^{N} (M^{q-1})_{ij}} \tag{53}$$

where q is a positive integer. Note that $(M^0)_{ij} = \delta_{ij}$ and for $q = 1$ we recover Eq. (52).

4.2.3 *The dispersal matrix as a ERM*

When developing a general theory of persistence in fragmented landscapes, we are faced with the problem that no two landscapes are alike. The situation is reminiscent of complex network theory, in which any two food webs, transportation, or gene-regulation networks are different, making it difficult to pinpoint the salient features of each system. In these cases, much progress has been made by contrasting empirical networks with those generated by simple models such as the Erdős-Rényi random graph or the Barabási-Albert model [5]. Our main goal is to propose and study a reference model for metapopulations dispersing in fragmented landscapes.

One could be tempted to simply take the Erdős-Rényi model and apply it to metapopulations [57]. However, this model lacks a fundamental feature of real dispersal networks: two patches close in space are more likely to exchange individuals than two that are far away. A more fruitful avenue is to take N patches, distribute them randomly in space, and connect any two patches that are closer than some threshold distance [58]. This defines a so-called Random Geometric Graph [53], for which it has been shown that the number of edges per node required to make the graph connected (i.e., the graph is composed of just one "piece") is much higher than that for Erdős-Rényi graphs, with the two converging for high-dimensional landscapes [59]. However, natural populations exist in low-dimensional environments. As such, Erdős-Rényi random graphs are inadequate descriptors of natural dispersal networks [57].

Random Geometric Graphs can be generalized even further. Instead of treating the connectedness of two patches in an "either/or" manner, we may think of it as a smooth function of distance. This function is what we refer to as the "dispersal kernel". Such networks have been introduced in the physics of disordered systems [50] and are called Euclidean Random Matrices. As explained above, a Random Geometric Graph is but one special case, in which the dispersal kernel is rectangular (Figure 12), yielding a dispersal rate of either 0 or 1, while generic Euclidean Random Matrices cover the broad spectrum of intermediate cases where dispersal rates decrease smoothly with distance (Figure 12).

Here we use Euclidean Random Matrices to study metapopulations dispersing in random fragmented landscapes. We derive a condition for metapopulation persistence analytically, highlighting that number of spatial dimensions, number of patches, shape of the dispersal kernel, and the variability in patch value are all key to determining persistence. We also show that a metapopulation can persist in two different regimes, localized and nonlocalized, giving rise to completely different responses to habitat loss. We finally obtain a counterintuitive result: if we arrange the patches in a perfect grid (as often considered

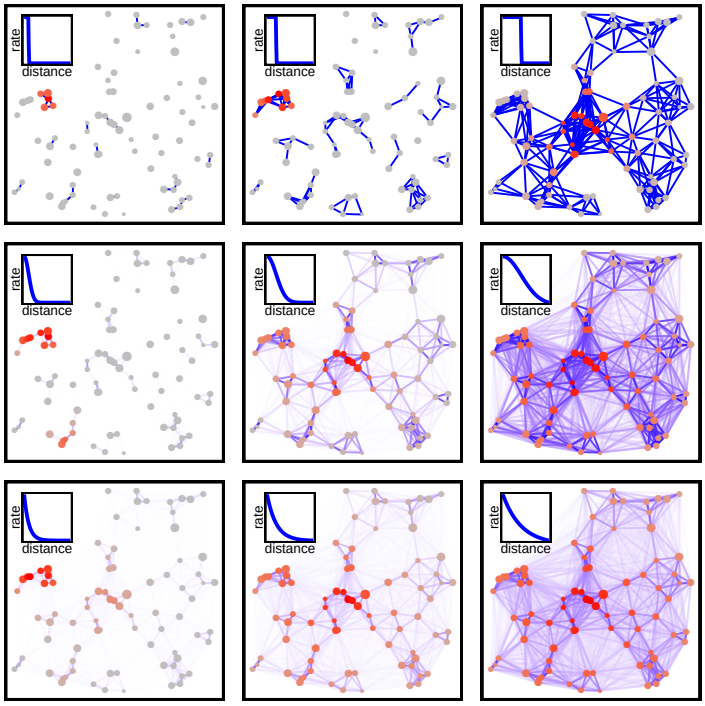


Figure 12: *Fragmented landscapes as networks. Different panels represent different dispersal kernels (top to bottom: Rectangular, Gaussian and Exponential—panel insets) for increasing dispersal lengths ξ (left to right). The size of the points stands for the patch value A_i, while the color measures the probability of occupancy p_i (gray $\rightarrow$ low probability, red $\rightarrow$ high probability). The color of the edges measures their strength (white to blue). In all the panels $\delta = 0.9\lambda$, where λ is the metapopulation capacity.*

in the design of protected areas [60, 61, 62]), we obtain a lower likelihood of persistence compared to the case in which we distribute the patches randomly.

4.3 METAPOPULATION CAPACITY IN FRAGMENTED LANDSCAPES

Knowing the number of dimensions (d) and size (L) of the landscape, the number of patches N, the dispersal kernel $f(|x_i - x_j|/\xi)$, and the distribution of the values of the patches (σ^2), we want to approximate λ, the metapopulation capacity.

Take the patch values to be one for all patches, and assume a rectangular kernel (Random Geometric Graph). Then, the average row sum is simply the average number of neighbors each patch has (the average degree of the network). When using another kernel (Euclidean Random Matrix), the row sum can be interpreted as the average number of "effective neighbors", n_e, meaning that patches that are closer contribute more to the sum than those that are far away. As such, when all patches have value one, $\lambda \geqslant n_e$. When the patch values

are sampled from a distribution with mean one and variance σ^2, assuming N large, we obtain $\lambda \geqslant n_e(1+\sigma^2)$ (see next sections). This provides a conservative criterion for metapopulation persistence:

$$n_e(1+\sigma^2) > \delta \, . \tag{54}$$

Next, we want to approximate n_e for different parameterizations. In fact, n_e is influenced by the dispersal kernel, the dispersal length, the number of dimensions, the number of patches, and the size of the landscape. The formula reads

$$n_e \approx \frac{N}{L^d} \, G_f(d) \xi^d \, , \tag{55}$$

(see section 4.3.1), where the first term measures the patch density, obtained dividing the number of patches N by the volume L^d (area if $d = 2$, or length if $d = 1$), while the second term $\left(G_f(d)\xi^d\right)$ measures the typical volume accessible via dispersal from a patch. Their product n_e represents therefore the typical number of patches accessible via dispersal. The numerical factor $G_f(d)$ depends on the functional form of the dispersal kernel and on d. For example, $G_f(d) = (2\pi)^{d/2}$ for the Gaussian kernel (see next sections).

Figure 13 shows that λ depends on all the factors that are needed to estimate n_e: the number of patches N, the kernel f, the dispersal length ξ, and the number of dimensions d. Moreover, λ depends on the variance in the value of the patches, σ^2. However, when plotting λ versus $n_e(1+\sigma^2)$, all curves collapse into the same one, meaning that two very different fragmented landscapes, having the same value of $n_e(1+\sigma^2)$, have approximately the same metapopulation capacity.

4.3.1 *Lower bound for metapopulation capacity*

Whether a metapopulation can persist depends on the metapopulation capacity λ, which in turn depends on the dispersal kernel f, the number of spatial dimensions d, and the density of suitable habitat patches. Using the Euclidean Random Matrices introduced in section 4.2.1, in this section we derive the metapopulation persistence criterion reported in equation 55.

In this section we examine the case when all patches have equal value $A_i = 1$, and therefore the matrix M defined by Eq. (48) is a Euclidean Random Matrix, with diagonal entries set to zero. Therefore the lower bound Eq. (52) can be used to approximate its leading eigenvalue. This bound is expressed as a double sum over the indices i and j which, in our case, means summing over the randomly distributed points in space:

$$\lambda \geqslant \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} M_{ij} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j \neq i}^{N} f\left(\frac{|x_i - x_j|}{\xi}\right) \, . \tag{56}$$

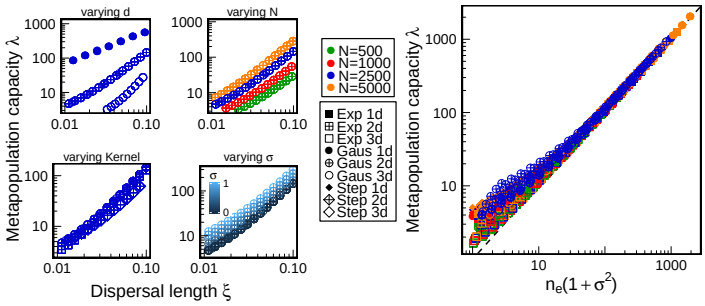


Figure 13: *Dependence of the metapopulation capacity on different factors. The four panels on the left show the metapopulation capacity* λ *vs. the dispersal length* ξ*.* λ *depends on the number of dimensions* d *(*1*,* 2*,* 3*), the number of patches* N *(*500*,* 1000*,* 2500*,* 5000*), the dispersal kernel* f *(Gaussian, Exponential and Rectangular), and the heterogeneity of patch value* σ *(varied between* 0 *and* 1*). Different symbols correspond to different combinations of* f *and* d*, while different colors correspond to different values of* N*. The color scale from blue to light blue refers to different values of* σ*. The right panel shows the collapsing of the various curves into a single one. The dependence on all the parameters (namely* d*,* N*,* ξ*,* f *and* σ*) is amalgamated in* $n_e(1+\sigma^2)$*, where* n_e *is a simple function of all the parameters (computed via Monte Carlo integration). The region of largest discrepancy (small* $n_e(1+\sigma^2)$*) corresponds to the localized case, where only few patches contribute to the persistence of the metapopulation.*

As discussed in Section 4.2.3, the expression above can be seen as the average row sum of matrix M. When M is a Random Geometric Graph (Section 4.2.1), each entry M_{ij} is either equal to 1 if patch i can be reached from patch j, or to 0 otherwise. Then, the sum of all entries in row i is the number of patches that are reachable from the ith patch. In other words, it is the number of neighbors of patch i. The average row sum is then the average number of neighbors of the patches over the landscape.

For other dispersal kernels, we do not have such a clear distinction between patches that are reachable and those that are not. Still, the row sum of row i can be thought of as the *effective number of neighbors* of patch i, because patches that are more likely to be reached from i contribute more to the sum than those that are difficult to reach. Therefore, we interpret the average row sum of matrix M as the average effective number of neighbors, denoted n_e.

When the number of patches is large, i.e., $N \gg 1$, we can approximate a summation over random points in a finite region Ω by a continuous integral:

$$\frac{1}{N}\sum_{i=1}^{N} g(x_i) \xrightarrow{N\to\infty} \frac{1}{V_\Omega}\int g(x)\,dx, \tag{57}$$

where g is an arbitrary function, V_Ω is the volume of the region Ω (or area for $d = 2$, length for $d = 1$), the integral runs over all of Ω, and

it is understood that "dx" refers to integration for all d components of x. Hence, for $N \to \infty$

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} f\left(\frac{|x_i - x_j|}{\xi}\right) \xrightarrow{N\to\infty} \frac{1}{V_\Omega^2}\iint f\left(\frac{|x-y|}{\xi}\right)\, dx\, dy. \tag{58}$$

The effective number of neighbors n_e can therefore be written, using the right hand side of Eq. (56), as

$$n_e = \frac{\rho}{V_\Omega}\iint f\left(\frac{|x-y|}{\xi}\right)\, dx\, dy\,, \tag{59}$$

where $\rho = N/V_\Omega$ is the density of patches over the landscape when $N \gg 1$. Then, from Eq. (56), we have $\lambda \geqslant n_e$. Since λ must be greater than δ for metapopulation persistence, this is a conservative estimate for the persistence condition: the metapopulation is expected to persist whenever the effective number of neighbors, estimated by the above integral, is larger than δ.

If we assume that the region Ω is the whole d-dimensional Euclidean space, the integration becomes especially simple. More precisely, we take the limit $N \to \infty$ and $V_\Omega \to \infty$ such that $\rho = N/V_\Omega$ remains finite. We then perform the change of variables $z = x - y$ to obtain

$$n_e = \rho\int f\left(\frac{|z|}{\xi}\right)\, dz, \tag{60}$$

a much simpler integral than in Eq. (59).

Given the dispersal kernel, we can now derive the persistence condition. We illustrate how this is done using three different kernel forms assuming Ω is the whole d-dimensional Euclidean space. The same calculation can be done for arbitrary other kernels in an analogous way.

- *Exponential kernel.* This kernel is given by

$$f\left(\frac{|x_i - x_j|}{\xi}\right) = \exp\left(-\frac{|x_i - x_j|}{\xi}\right), \tag{61}$$

  where ξ is the dispersal distance. Let the points of Ω be denoted by x with $x_{(k)}$ being its kth coordinate ($k = 1 \ldots d$). Since we are integrating over the whole space, we can use Eq. (60), yielding

$$\begin{aligned} n_e =& \rho\int \exp\left(-\frac{\sqrt{\sum_{k=1}^{d} x_{(k)}^2}}{\xi}\right)\, dx_{(1)}\cdots dx_{(d)} \\ =& \frac{2\pi^{d/2}\Gamma(d)}{\Gamma(d/2)}\rho\xi^d\,, \end{aligned} \tag{62}$$

  where Γ is the gamma function.

- *Gaussian kernel.* This is given by

$$f\left(\frac{|x_i - x_j|}{\xi}\right) = \exp\left(-\frac{|x_i - x_j|^2}{2\xi^2}\right) \tag{63}$$

with dispersal distance ξ. Again, we apply Eq. (60) to obtain

$$\begin{aligned} n_e =& \rho \int \exp\left(-\frac{\sum_{k=1}^{d} x_{(k)}^2}{2\xi^2}\right) dx_{(1)} \cdots dx_{(d)} \\ =& (2\pi)^{d/2} \rho \xi^d \,. \end{aligned} \tag{64}$$

- *Rectangular kernel.* This kernel is the basis for Random Geometric Graphs [53]. It reads

$$f\left(\frac{|x_i - x_j|}{\xi}\right) = \begin{cases} 1, & |x_i - x_j| \leqslant \xi, \\ 0, & |x_i - x_j| > \xi, \end{cases} \tag{65}$$

with ξ again being the dispersal distance. Integrating over the whole Euclidean space, we are actually obtaining the volume of a d-dimensional sphere of radius ξ. The effective number of neighbors is therefore given by

$$n_e = \frac{\pi^{d/2}}{\Gamma(d/2+1)} \rho \xi^d \,. \tag{66}$$

Since the effective number of neighbors n_e is an approximation for the average row sum of M, and this quantity itself is a lower bound for the metapopulation capacity λ, we used simulations to see how well the approximation works in practice. Figure 13 shows that when plotted against n_e the eigenvalues obtained with different parameterizations collapse. Figure 14 shows λ against n_e for various dispersal kernels and number of dimensions d. The region Ω is taken to be a d-dimensional cube of unit volume, and the dispersal distance ξ is varied between 0 and 0.2. The effective number of neighbors n_e is obtained by evaluating the integral Eq. (59) numerically within the unit cube, while λ is numerically calculated as the actual leading eigenvalue of the Euclidean Random Matrix M.

Figure 14 shows that, unless the effective number of neighbors is very low, the approximation is very good. Even when n_e is small, the approximation errs on the conservative side: the actual leading eigenvalue is in fact larger than predicted, leading to a greater likelihood of persistence than n_e predicts.

To better show the difference between the two quantities, Figure 15 shows the value $(\lambda - n_e)/n_e$ against n_e, i.e., the relative deviation of λ from n_e. Interestingly, this relative deviation seems to satisfy a simple power-law relationship

$$\frac{\lambda - n_e}{n_e} \approx Q_f(d) n_e^{-1/4} \,, \tag{67}$$

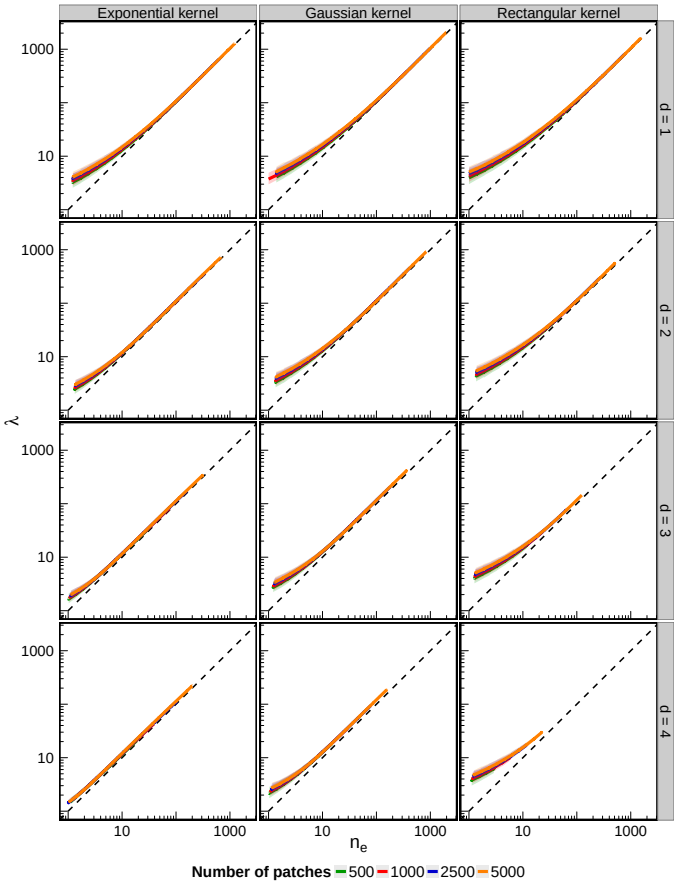


Figure 14: Approximating the metapopulation capacity λ using the effective number of neighbors n_e. There are 4×3 insets; the rows correspond to the dimension d, and the columns to different dispersal kernels. Each curve is obtained from a set of simulations with given d, kernel, number of patches (colors), and dispersal distance ξ, (varied between 0 and 0.2). For each parameter combination, we performed 1000 simulations. In each simulation patches of identical value are scattered uniformly in the $d-$dimensional unit cube, M is calculated using the parameters, λ is obtained by numerically finding the leading eigenvalue of M, and n_e is obtained by performing the integral in Eq. (59). The approximation is conservative when n_e is small, and is very accurate for larger values.

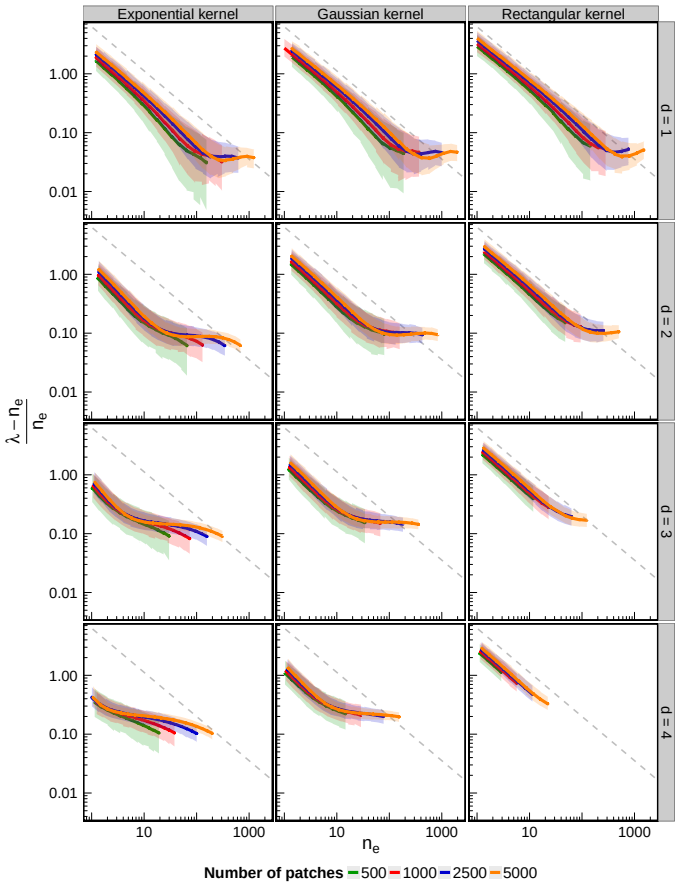


Figure 15: Relative deviation of the metapopulation capacity λ from the effective number of neighbors n_e. Zero deviation means the effective number of neighbors provides a perfect estimate of the metapopulation capacity. Methods and notation as in Figure 14. The gray dashed lines indicate a power-law relationship with exponent $-1/4$. The actual results approximate this ideal curve very well, suggesting the simple rule $(\lambda - n_e)/n_e \approx Q_f(d) n_e^{-1/4}$. The factor Q depends on the number of dimensions d and the general form of the dispersal kernel, but not on the dispersal distance ξ or the number of patches N. Rearranging, we get the improved approximation $\lambda \approx n_e + Q_f(d) n_e^{3/4}$ for the metapopulation capacity in place of $\lambda \approx n_e$. Efficient practical usage of this improved formula would of course require a knowledge of the function Q.

where the exponent $-1/4$ is universal being independent of the dispersal kernel and the dimensionality d. Despite the fact that we do not have an analytic explanation for this empirical result, Figure 15 shows that this seems to be a very robust pattern, with potential implications for further developments in the theory of Euclidean Random Matrices.

4.3.2 *Localization*

In this section we discuss two further features of metapopulation structure over random fragmented landscapes. The first is variance in patch importance, meaning that some habitat patches are more important for the metapopulation's persistence than others. The second is spatial localization: patches with high occupancy probabilities p_i tend to be close together in space.

The components of the leading eigenvector of M provide a measure of the importance of a patch for persistence. It can be shown [63] that the relative change in the metapopulation persistence due to the removal of patch i is approximately equal to w_i^2, where w_i is the ith component of the leading eigenvector. As we have shown above, the eigenvector is also related, in the limit $\delta \approx \lambda$, to the stationary patch occupancy probabilities p_i. High variance in the eigenvector components therefore corresponds to heterogeneity in patch importance.

To study the variance in patch importance quantitatively, we use a metric applicable to any nonzero vector w with nonnegative components that is normalized to have length 1,

$$\sum_{i=1}^{N} w_i^2 = 1, \tag{68}$$

where w_i is the ith component of the vector w. The *inverse participation ratio* (IPR [64]) is defined as

$$\mathrm{IPR} = \sum_{i=1}^{N} w_i^4 \,, \tag{69}$$

and measures the heterogeneity of eigenvector components. Here we rescale this metric so that its values fall between 0 and 1:

$$\Psi = \frac{N \times \mathrm{IPR} - 1}{N - 1} = \frac{N \sum_{i=1}^{N} w_i^4 - 1}{N - 1}. \tag{70}$$

When patches are arranged in a perfect grid, we do not expect any variance in the w_i (with the exception of the possible small influence of boundary effects). To measure how patch occupancies changes for disordered systems, we first perturbed a regular grid of patches, by adding to each position a random vector whose component are drawn from a uniform distribution between $-\eta$ and η. In Figure 16 we show

the scaled inverse participation ratio Ψ of the normalized leading eigenvector w against the metapopulation capacity λ for various values of the perturbation size η. We see that increasing randomness leads to higher variation in all cases.

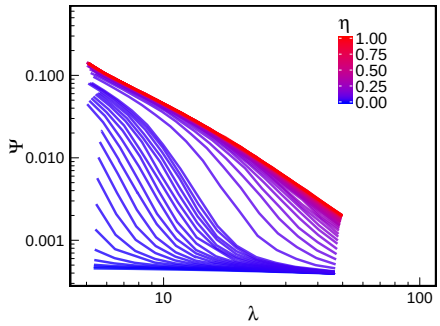


Figure 16: The effect of a grid structure versus randomly distributed patches on the variance in patch occupancies. The scaled inverse participation ratio Ψ of the normalized leading eigenvector w, defined by Eq. (70), is plotted against the metapopulation capacity λ. Different colors represent different values of η, which quantifies the level perturbation of the perfect grid ($\eta = 0$ correspond to a grid, $\eta = 1$ to a random distribution of points). Ψ increases with the perturbation size (increasing randomness increases Ψ), and decreases with the metapopulation capacity. Note that, since our formula for Ψ depends on the approximation Eq. (52), i.e., the $\lambda \approx \delta$ limit, the metapopulation is still close to extinction even for large values of the metapopulation capacity.

This simulation sheds light on the cause for the discrepancy between λ and n_e observed in Figures 14 and 15. We have shown that n_e approximates λ very well, except when n_e is very small. The relationship between the two quantities is of course not expected to be perfect since the effective number of neighbors only provides a lower bound for λ, as discussed in Section 4.3.1. The only case when the lower bound actually holds with strict equality is when w is completely uniform, i.e., is proportional to the vector $(1, 1, \ldots)$. But this implies that $\Psi = 0$, corresponding to the lack of any variation in patch occupancies. Any deviation from this situation leads to more variance and therefore a worse approximation due to Eq. (56). Therefore, any discrepancy between λ and n_e is strictly due to this variance.

Instead of relating Ψ to the degree of randomness in the distribution of patches across the landscape, one may also ask how it relates to the metapopulation capacity itself. To answer this question, we generated 1000 matrices M for each combination of the following parameters: N (which could take on the values 500, 1000, 2500, and 5000), d (between 1 and 4), kernel (Exponential, Gaussian, and Rectangular), and dispersal distance ξ (taking on 84 possible values between 0 and 0.2). In each case we determined Ψ and λ for each M. The scaled in-

verse participation ratio Ψ was then calculated. Figure 17 shows the results. We can see that Ψ is inversely related to λ in all cases: a larger metapopulation capacity leads to less variance in patch occupancies. Though this result might seem self-evident, we have assumed $\lambda \approx \delta$, and therefore high values of the metapopulation capacity do not in any way imply strong metapopulation persistence—on the contrary, regardless of the magnitude of λ, the metapopulation is very close to the extinction threshold.

Having established that there is variation in the importance of patches for metapopulation persistence, we can also ask whether patches with similar importance tend to form spatial clusters, i.e. whether patches of similar importance are close together on the landscape. To answer this question, we measured the correlation between closeness and importance using Moran's index [65]. Independent of n_e, this index turns out to be positive and significant (p-value always less than 10^{-6}), i.e. close patches have similar importance. This positive correlation is expected, since in terms of network theory, the importance of a node is measured by the eigenvector's components (the so-called eigenvector centrality). It directly follows from the definition of eigenvector centrality (important nodes are nodes connected with important nodes) that strongly connected patches have similar importance. But the patches that are going to be strongly connected are the ones close in space, since the strength of connection depends only on the distance between patches in our Euclidean Random Matrix framework. This argument already implies a positive Moran's index—which we indeed do find.

4.3.3 *Effect of patch heterogeneity*

So far we have assumed that all patches have the same value, i.e., in Eq. 48 the A_i are all equal. In fact, we have assumed $A_i = 1$; however, for any constant patch value $A_i = a$ one can rescale the equilibrium equation Eq. 47 by dividing both sides by a^2 and treating δ/a^2 as a new effective δ. Here we introduce variable patch values.

To explore their effects on metapopulation structure and persistence, we randomized the A_i by drawing them independently from a scaled beta distribution with mean 1 and variance σ^2 (again, without loss of generality).

We can use Eq. (53) with $q = 2$ to derive an improved approximation for the metapopulation capacity when the patch values are variable:

$$\lambda \geqslant \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \left(M^2\right)_{ij}}{\sum_{i=1}^{N} \sum_{j=1}^{N} M_{ij}}, \tag{71}$$

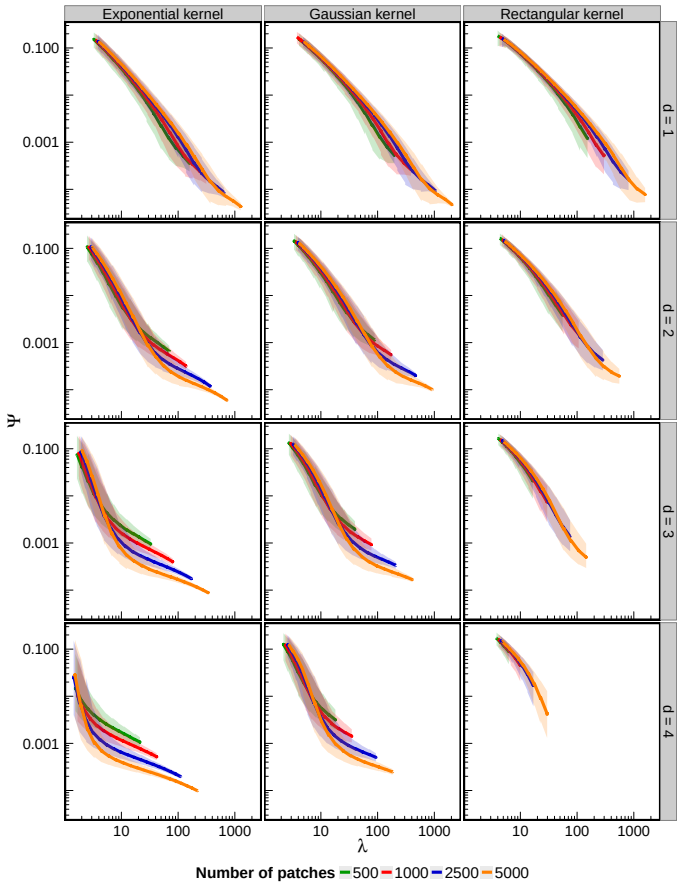


Figure 17: The scaled inverse participation ratio Ψ as a function of the metapopulation capacity in the $\lambda \approx \delta$ limit. Methods and notation as in Figure 14. Ψ diminishes with an increasing metapopulation capacity in all cases.

or, writing out the matrix multiplication in the numerator,

$$\lambda \geqslant \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} M_{ik}M_{kj}}{\sum_{i=1}^{N}\sum_{j=1}^{N} M_{ij}}. \tag{72}$$

Substituting in Eq. (48) and using the notation $f_{ij} = f\left(|x_i - x_j|/\xi\right)$ when $i \neq j$ and 0 for $i = j$,

$$\lambda \geqslant \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} A_i A_j A_k^2 f_{ik} f_{kj}}{\sum_{i=1}^{N}\sum_{j=1}^{N} A_i A_j f_{ij}}, \tag{73}$$

where the matrix f is a Euclidean Random Matrix.

The matrix f has zeros on the diagonal, and therefore the sum in the numerator always yields zero whenever $k = i$ or $k = j$. Since f is a function of independently drawn random points in space, and the patch values are also independently drawn, the numerator can be treated as an averaging over the patches via the index k. Let us define the quantity

$$L_k = \sum_{i=1}^{N}\sum_{j=1}^{N} A_i A_j f_{ik} f_{kj}. \tag{74}$$

We then have

$$\lambda \geqslant \frac{\sum_{k=1}^{N} A_k^2 L_k}{\sum_{i=1}^{N}\sum_{j=1}^{N} A_i A_j f_{ij}}, \tag{75}$$

or

$$\lambda \geqslant \frac{N\overline{A^2 L}}{\sum_{i=1}^{N}\sum_{j=1}^{N} A_i A_j f_{ij}}, \tag{76}$$

where the bar denotes averaging over the patches. Due to independence, we have $\overline{A^2 L} = \overline{A^2}\,\overline{L}$. The quantity $\overline{L}$ can be written

$$\begin{aligned}\overline{L} =& \frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{N}\sum_{j=1}^{N} A_i A_j f_{ik} f_{kj} = \frac{1}{N}\sum_{k=1}^{N}\left(\sum_{i=1}^{N} A_i f_{ik}\right)\left(\sum_{j=1}^{N} A_j f_{kj}\right)\\ =& \frac{1}{N}\sum_{k=1}^{N}\left(\overline{A}\sum_{i=1}^{N} f_{ik}\right)\left(\overline{A}\sum_{j=1}^{N} f_{kj}\right) = \frac{(\overline{A})^2}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(f^2\right)_{ij},\end{aligned} \tag{77}$$

where we used the independence of A and f again. The denominator of Eq. (76) can be written

$$\begin{aligned}\sum_{i=1}^{N}\sum_{j=1}^{N} A_i A_j f_{ij} =& \sum_{i=1}^{N} A_i \sum_{j=1}^{N} A_j f_{ij} = \overline{A}\sum_{i=1}^{N}\sum_{j=1}^{N} A_j f_{ij}\\ =& (\overline{A})^2 \sum_{i=1}^{N}\sum_{j=1}^{N} f_{ij}.\end{aligned} \tag{78}$$

Substituting Eqs. (77) and (78), we obtain

$$\lambda \geqslant \overline{A^2} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \left(f^2\right)_{ij}}{\sum_{i=1}^{N} \sum_{j=1}^{N} f_{ij}}. \tag{79}$$

We discussed in Section 4.2.3 that, while both Eq. (52) and Eq. (53) (with $q = 2$) approximate λ, the latter is always greater than or equal to the former:

$$\frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \left(f^2\right)_{ij}}{\sum_{i=1}^{N} \sum_{j=1}^{N} f_{ij}} \geqslant \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} f_{ij}}{N}. \tag{80}$$

Therefore we may substitute this into Eq. (79) and the inequality will still hold with

$$\lambda \geqslant \frac{\overline{A^2}}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} f_{ij}. \tag{81}$$

But the above sum divided by N is simply the effective number of neighbors n_e, as discussed in Section 4.3.1. We therefore have

$$\lambda \geqslant \overline{A^2}\, n_e. \tag{82}$$

Since we are drawing the patch values independently from a distribution with mean 1 and variance σ^2, and $\overline{A^2} = (\overline{A})^2 + \mathrm{Var}(A)$, we have $\overline{A^2} = 1 + \sigma^2$. We therefore obtain

$$\lambda \geqslant (1 + \sigma^2) n_e. \tag{83}$$

We tested how the original $\lambda \geqslant n_e$ and the new $\lambda \geqslant (1 + \sigma^2) n_e$ approximations perform as a function of the patch variability σ. The left panel of Figure 18 shows the former, the right panel the latter approximation. On the left panel, for small values of σ we see exactly the same situation found in Figure 14, but for large values of σ the effective number of neighbors starts to severely underestimate the metapopulation capacity. On the right panel however, we see that as long as $(1 + \sigma^2) n_e$ is not very small, this quantity approximates the metapopulation capacity very well even for large patch variability.

4.4 CONCLUSIONS

By modeling fragmented landscapes as networks [66, 67] in which the nodes are patches and the weighted edges represent dispersal, we have shown that metapopulation persistence can be studied analytically for the case in which patches are randomly distributed in the landscape, and patch values are independently sampled from a distribution.

The derivation highlights that a few key quantities determine the metapopulation capacity: the density of the patches, with denser landscapes yielding a higher probability of persistence; the shape of the

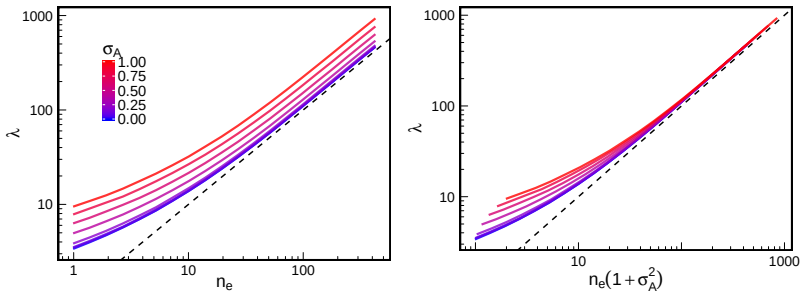


Figure 18: The effect of nonconstant patch values on the approximation $\lambda \approx n_e$. On the left panel we plot the metapopulation capacity (obtained by computing λ numerically) against the effective number of neighbors n_e, calculated via the integral in Eq. (59). The patch values were drawn from a rescaled beta distribution with mean 1 and standard deviation σ. The matrices M were generated with $N = 2500$, $d = 2$, a Gaussian dispersal kernel, the dispersal distance ξ, varying between 0 and 0.2, and σ varying between 0 and 1. For every parameter combination, 1000 simulations were performed. As seen on the plot, the greater the variability in patch value, the more n_e underestimates λ. The right panel shows the exact same information, except the metapopulation capacity is plotted against $n_e(1+\sigma^2)$ instead of just n_e (section 4.3.1). The approximation is significantly improved by including the term $(1+\sigma^2)$.

dispersal kernel; the number of dimensions; and the variability in patch value, with higher variance being beneficial.

Our analysis provides a null model for metapopulation persistence. For a given empirical landscape, in which the patches positions are not necessarily described by a uniform distribution and patch values are not independent of patch position[68, 57], the effect of these features on persistence can be disentangled from the effects of other factors by contrasting the metapopulation capacity of the empirical landscape with what expected according to our framework.

Interestingly, we found that disorder and variation are beneficial for persistence. First, the variance in patch values has a very strong positive effect on the metapopulation capacity, meaning that highly heterogeneous patch values yield higher λ that homogeneous ones. Second, we found that more "disordered" arrangements of the patches increase both the metapopulation capacity and the expected proportion of occupied patches. This is especially relevant when metapopulations are close to extinction, as localization becomes key for maintaining the metapopulation viable, albeit in a spatially confined region. Note that these analytic results shed light on previous simulations suggesting that unequal spacing between the patches is beneficial for populations close to extinction[62], and are consistent with what found when considering stochastic dynamics[69].

Although we have examined the case of randomly distributed patches using a uniform distribution, the same approach holds when the distribution is not uniform[68]. For example, riparian plants in a landscape crossed by a river will be concentrated in the vicinity of the water, leading to a non-uniform distribution. Fortunately, in the limit of many patches, any distribution can be taken into account, by evaluating analytically or numerically the integral defining n_e in the general case (see eq. 67). Similarly, we accounted for an integer number of dimensions, but a fractional d (e.g., in a fractal-like river basin) would not alter the framework. Finally, we assumed that species can disperse equally in all directions, with no preference. When there is a clear preferential direction of dispersal (e.g., wind-dispersal of seeds, or fish larvae dispersing in a river), the theory of Euclidean Random Matrices can be extended to account for this lack of symmetry[70].

The derivation of a criterion for metapopulation persistence bears a striking resemblance to the derivation of stability criteria for large ecological communities[25, 32], as in both cases the use of random matrix theory led to the identification of the few basic parameters responsible for the large-scale behavior of the systems. The advantage of this approach is that, once the modeling of the matrices is in place, the derivation of the results requires only elementary algebra. Random matrix theory is currently experiencing an impressive growth[71], greatly expanding the potential for biological applications.

Part II

CRITICALITY IN BIOLOGICAL INTERACTIONS

Perhaps our ultimate understanding of scientific topics is measured in terms of our ability to generate metaphoric pictures of what is going on.
Maybe understanding is coming up with metaphoric pictures.

— Per Bak

5 CRITICALITY IN LIVING SYSTEMS

Facts have to be discovered by observation,
not by reasoning .

— Bertrand Russell

5.1 BRIEF PRIMER ON CRITICAL PHENOMENA

Critical phenomena are well understood in physical systems. The great lesson learned from statistical physics is that, even though the elementary constituents can be as simple as spheres or spins sitting on a lattice with pair-wise interactions, criticality and scale invariance emerge as the collective behavior of a many-body system with its characteristics depending only on just a few essential attributes such as the dimensionality of the system and symmetries of the problem. Remarkably, this universality results in the critical behavior of an Ising model with nearest neighbor interactions on a cubic lattice being identical to that of a liquid-vapor system at its critical point. Likewise, the critical behavior of a binary alloy that is about to order is the very same as the two other cases. This is because all these systems are three dimensional and have the same "up-down" symmetry. The underlying details: the fact that spins sit on an idealized lattice, the chemistry of the liquid, or the atomic interactions in an alloy are irrelevant in determining the critical behavior.

A classical Ising spin can point up or down. An interaction between neighboring spins results in favoring a parallel relative orientation over an antiparallel one. This is captured through an interaction energy that is lower and thus more favorable when neighboring spins are parallel compared to when they are antiparallel. Note that our description of the model favors parallel over antiparallel but up and down are treated symmetrically. The ground state or the lowest energy state of such an Ising system is one in which all spins are parallel and, by necessity, are all up or all down. This choice between up and down breaks the up-down symmetry spontaneously. The advantage of lowering the energy through a mostly parallel alignment and thus breaking the symmetry, favored at low temperatures, competes with the tendency to increase the entropy at high temperatures through a restoration of the symmetry and having roughly equal numbers of up and down spins. While there are just two states with perfectly parallel spins –all up or all down– there are many states and therefore a higher entropy when approximately half the spins are up and the other half are down. The magnetization of such an Ising system is pro-

portional to a suitably normalized imbalance between the numbers of up and down spins. It provides a measure of the ordering and is zero at high temperatures and is 1 at zero temperature. On lowering the temperature from a very high value, the magnetization remains zero until a critical temperature is reached at which point the magnetization rises continuously and becomes non-zero. The critical point is then a special temperature at which there is an onset of a non-zero magnetization and the up-down symmetry is spontaneously broken. There are two phases that emerge: at any temperature higher than T_c, the magnetization is zero for an infinite sized system whereas, below T_c, the magnetization is non-zero. The critical point, which separates these two phases, is obtained by tuning the temperature just right to its critical value. At a critical point, there are domains of up and down spins of all sizes thoroughly interspersed among one another. Scale invariance occurs because there is no dominant size scale associated with these domains and power law correlations between spin orientations are observed.

In inanimate matter is well understood how to reach the critical point, in terms of control parameters, and what are the relevant properties that allow to identify a state of the matter as critical. When one considers biological systems the situation is radically different. Living systems (from flocks [74] to ecosystems themselves [75]) are inherently out of equilibrium and there is not yet any fundamental understanding of the control parameters of those systems. In this context the critical point could be an odd defined property. Despite this fact, there are a lot of examples where biological systems has properties that remind of the critical behaviors that we are familiar of. In the next section we will introduce and discuss some of those possible examples.

5.2 ARE BIOLOGICAL SYSTEMS POISED AT CRITICALITY?

Most of the complex biological phenomena emerge from interactions among many constituents [76]: the interactions among many genes determine how cells look like and behave, the interactions among neurons determine how we think and have memories, the interactions among species determine the fate of an ecosystem. Despite their constituent role, we have not a fundamental understanding of the principles behind those biological interactions.

One possible approach to study interactions is via an inverse problem. This is a data-driven way to build an effective Hamiltonian and give a statistical mechanics description of the system of interest. This approach is based on maximum entropy principle (maxent). In a nutshell, one choose some observables (constraints) from the data and find the probability distribution that maximize the entropy conditioned to reproduce the constraints as averages. More precisely, one

has a series of data $\{s\}$ and some functions of the data $O^\mu(\{s\})$ (observables). The goal is to find a probability distribution $P(\{s\})$ that reproduces the data. In the maxent approach this is achieved by maximizing the entropy of the probability

$$S[P] = -\sum_{\{s\}} P(\{s\}) \log P(\{s\}) \,, \tag{84}$$

constrained to reproduce the observables

$$O^\mu_{data} = \langle O^\mu(\{s\})\rangle = \sum_{\{s\}} P(\{s\}) O^\mu(\{s\}) \,, \tag{85}$$

where O^μ_{data} is the value of $O^\mu(\{s\})$ computed from the data. The distribution that maximizes the entropy constrained to the observables has the form

$$P(\{s\}) = \frac{\exp(\sum_\mu g_\mu O^\mu(\{s\}))}{Z} \,, \tag{86}$$

where Z is a normalization constant, while g_μ are the Lagrange multipliers that are fixed to satisfy the constraints of equation 85. The probability in equation 86 has the same form of the probability of statistical mechanics system with Hamiltonian equal to $-\sum_\mu g_\mu O^\mu(\{s\})$ (up to a factor β).

Given a dataset and an arbitrary set of constraints one can obtain a full probability distribution using maxent. If one chooses the right constraints, the probability distribution obtained from equation 86 well reproduces the data and not only the imposed constraints. In this way one has a parametrization for the system of interest, where the tunable parameters are g^μ and the empirical system corresponds to the value of the parameters reproducing the constraints. It is important to stress that this is a very important assumption, as these tunable parameters of the model g^μ do not need to correspond to any realistic parameter tunable in an experiment.

Maxent analysis has been performed on many biological systems (from neurons [77] to ecosystems [78, 79]). One question that naturally arises when this analysis is performed is the position of the empirical system in the phase space obtained from the system of equation 86. Many biological system turn out to have parameters corresponding to the critical ones. Here below we will consider explicitly some examples.

5.2.1 *Network of neurons*

In the nervous system of all the animals, neurons communicate with one another through discrete pulses of electrical activity known as spikes. One can therefore imagine to discretize the electric signal measured from a group of neurons. By looking at the temporal series divided in small windows of time, the continuous temporal series of the

neuron i is transformed into a series of discrete variables σ_i^t taking values 1 if the neuron is spiking in that window and 0 if it is not.

In order to apply maxent one has to choose constraints and a natural choice in this case is to fix the average activity and the correlation between neurons, that result in the following probability

$$P(\{s\}) = \frac{\exp(\sum_i h_i \sigma_i + \sum_{ij} J_{ij} \sigma_i \sigma_j)}{Z} , \tag{87}$$

which corresponds to the Gibbs distribution of an Ising model. The parameters h_i are related to the constraints of the average activity, while J_{ij} can be interpreted as an effective interactions between neurons.

Do the inferred interactions have any peculiar property? In ref. [76] it was shown that the values of the parameters seem to be poised at the critical point. In order to do that they introduced an effective inverse temperature β and considered the probability

$$P(\{s\}|\beta) = \frac{\exp(\beta(\sum_i h_i \sigma_i + \sum_{ij} J_{ij} \sigma_i \sigma_j))}{Z} . \tag{88}$$

When $\beta = 1$ one recovers the model fitted from the data, while tuning that parameters one effectively changes the value of parameters. The authors studied the "specific heat" of the model, defined as the variance of the Hamiltonian (in this case the argument of the exponential), in relation with β. In an equilibrium system the specific heat diverges at the critical point in the thermodynamic limit. If the system is finite but large enough we know that the maximum of the specific heat is located closer and closer to the critical point as the size of the system increases. Retinal ganglion neurons show a peak at $\beta = 1$, indicating that the system is possibly close to a critical point [76].

The one described above is just one example of observables suggesting that empirical neural networks operate close to the critical point, and many other observables suggest, or at least are compatible, with a dynamics at the critical point [80, 81, 82].

5.2.2 *Animal collective behavior*

Flocks of birds are a remarkable example of collective behavior. The flock acts as a unique organism, responding in a very coordinated way to external perturbations.

In recent years both experimental and theoretical effort have increased our understanding of flocks and swarms [83, 84, 85, 86, 87]. Many evidences suggest that the interaction between individuals is somewhat tuned to a critical point [88, 89]. Here we briefly review some of them.

Maximum entropy has been applied to flocks of birds to study the orientation of velocities of individuals. When the correlation between

the velocity of birds is used as a constraint, the resulting model is equivalent to an Heisenberg model in 3 dimensions.

$$P(\{\underline{s}\}|\beta) = \frac{\exp(\sum_{ij} J_{ij}\underline{s}_i\underline{s}_j)}{Z}\,, \tag{89}$$

where $\underline{s}_i = \underline{v}_i/\|\underline{v}_i\|$, being $\underline{v}_i$ the velocity of individual i. The coefficient J_{ij} can be interpreted by using the correlation of orientation across individuals. The orientation shows a very strong polarization, i.e. $\|\sum_i \underline{s}_i\|/N$ is very close to one. This suggests that the parameter of the Heisenberg model are set in the ferromagnetic phase and the system has broken the symmetry. What is therefore relevant are the fluctuations around the polarized direction. Interestingly the correlations among these fluctuations display a long range correlation [85]. Empirical data shows that the correlation length increases with the system size [85], which is a signature of a critical point. Note that in this case the fact that the correlation between fluctuations is scale-free is not just a suggestive analogue of critical phenomena, but is predicted correctly by the model obtained via maxent.

Very similar results applies to swarm of midges [89, 87], that displays similar correlation, with a correlation length that grows with the system size. Moreover in this case, one can introduce a dynamical description, the Vicsek model [90], that well reproduces the data and predict a transition between a low-density disordered phase to a polarized ordered one. Empirical data lies on values of parameters corresponding to the maximum of the correlation length and the maximum of the susceptibility.

5.3 SCALE INVARIANCE AND CRITICALITY

Most of the arguments suggesting that living systems operate close to a critical point are based on scale invariance. While it is true that a (second order) critical point implies scale-invariance, the opposite is not generally true.

In many context data are indeed scale-free and a signature of this is the Zipf's law. Zipf's law was firstly introduced to describe the distribution of word in texts and it states that, if σ is a word and $P(\sigma)$ is its frequency, than

$$P(\sigma) \propto \frac{1}{r(\sigma)}\,, \tag{90}$$

where $r(\sigma)$ is the rank of the word σ. Zipf's law can be generalized to an arbitrary power-law relation between P and r and it has been observed in many context, from the distribution of gene family content in genomes [91, 92] to the sizes of cities [93], from the income distribution of companies [94] to earthquakes [95].

It has been proposed a connection between Zipf's law and equilibrium system at the critical point [76]. One can indeed show that an

equilibrium model at the critical point has a distribution of configurations following a power-law. Self-organized criticality [96, 72] is a different, yet similar, possible explanation behind the emergence of Zipf's law in many systems. In this case scale invariant phenomena emerge as properties of an attractor of a dynamical system, without fine-tuning of any parameter. Zipf's law has also been proposed to be originated by the interaction of the system with unknown hidden variables [97]. When the distribution of external variables is sufficiently broad, Zipf's laws emerges.

It is interestingly to note that the Zipf's law was introduced to describe the word distribution in text. Notably in this case Zipf's law is not an expression of any interesting interactions or critical properties in linguistics, but just a consequence of the simple fact that words are made of letters [98]. Random text, simply generated by assembling letters and spaces, show a Zipf's law in word count. Scale-invariance phenomena could be a consequence of other processes than criticality.

Criticality is interesting (mainly) because of fine-tuning. To sit at the border of a critical point and be able to observe critical phenomena one has to carefully tune the temperature or any other control parameter. In this context it is important to note that fine-tuning means that the region of critical temperatures is very small (and tends to zero in the thermodynamic limit) respect to the typical range we are used to. When we deal with complex systems we have no knowledge of what are the control parameters and their relation with the parameters that are really tunable in the system. That is to say that critical behavior can, at least in principle, correspond to a large range of parameters the system has access to and being effectively described by a model with parameters tuned at the critical point.

The other side of the coin is the richness of behaviors that appear at the critical point. Close to the critical point there are many more distinguishable models than far from it. If we think in terms of parameters, the region corresponding to critical phenomena is very small, while in terms of distinguishable models, the most of them are close to the critical point. In particular, one can show that the density of different models is proportional to the generalized susceptibility [99], that diverges at the critical point.

5.4 WHAT NEEDS TO BE DONE

As explained in previous sections many properties that are reminiscent of critical phenomena are observed in several biological systems. Whether the interactions between the fundamental constituents of these system are or not poised at the critical point, is still an open problem. In this section we speculate on the possible way to shed light on why these systems have these critical-like properties.

Experiments. Most of the system where critical-like behaviors are observed are non controllable systems. In some system that is practically impossible, one cannot tune the parameters governing the interactions between birds in flocks, while in others is more feasible, e.g. one can tune the interactions between neurons in a proper designed in-vitro experiment. Being able to tune some control parameters allows to check weather the original, natural, value of those parameters is more critical than the modified system.

Simple null model. Some of the critical-like properties could be in principle null properties, as in the case of Zipf's law and random text. Having proper null models able to describe the observed pattern does non necessarily mean that the system is out of the critical point, but give an idea of the effective parameters (that are non necessarily the same as the controllable one) governing the system.

In the next sections we describe a different approach to the problem. Instead of describing a specific system, we try to understand why a generic community of individuals, under very general assumption, could show critical-like properties.

6

INFORMATION-BASED FITNESS AND CRITICALITY

If the doors of perception were cleansed
everything would appear to man as it is,
infinite

— William Blake

6.1 INTRODUCTION

In this chapter we describe a very abstract framework trying to explain why many systems could opertate close to a critical point and how the critical point can be reached as a stable point of a proper dynamics.

As conjectured long ago, the capability to perform complex computations, which turns out to be the fingerprint of living systems, is enhanced in "machines" operating near a critical point [101, 102, 103], i.e. at the border between two distinct phases: a disordered phase, in which perturbations and noise propagate unboundedly –thereby corrupting information transmission and storage – and an ordered phase where changes are rapidly erased, hindering flexibility and plasticity. The marginal, critical, situation provides a delicate compromise between these two impractical tendencies, an excellent trade-off between reproducibility and flexibility [81, 104, 105] and, on larger time scales, between robustness and evolvability [106]. A specific example of this general framework are genetic regulatory networks [103, 107]. Cells ranging from those in complex organisms to single-celled microbes such as bacteria respond to signals in the environment by modifying the expression of their genes. Any given genetic regulatory network, formed by the genes (nodes) and their interactions (edges) [108]– can be tightly controlled to robustly converge to a fixed almost-deterministic attractor –i.e. a fixed "phenotype"– or it can be configured to be highly sensitive to tiny fluctuations in input signals, leading to many different attractors, i.e. to large phenotypic variability [109]. These two situations correspond to the ordered and disordered phases respectively. The optimal way for genetic regulatory networks to reconcile controllability and sensitivity to environmental cues is to operate somewhere in between the two limiting and impractical limits alluded to above [103] as has been confirmed in different experimental set-ups [110, 111, 112]. Still, it is not clear how such tuning to criticality comes about.

6.2 MODEL FRAMEWORK

Living systems operate in a highly variable and largely unpredictable external environment. In our framework, we describe living systems and the external environment in terms of probability distribution functions. Living systems need to modify their internal state to cope with external conditions. We represent an environmental signal "perceived" and processed by a living system as a string of N (binary) variables, $\mathbf{s} = (s_1, s_2, \ldots s_N)$. A specific environmental source is modeled by the probability distribution P_{src} that specify the frequency of each of the 2^N possible environmental signals. This distribution is assumed to depend on a set of parameters, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots)$, accounting for environmental variability. An individual living system or "agent" seeks to adapt itself to cope with the perceived stimuli/signals produced by a given environmental source. This is accomplished by changing its internal state, encapsulated in a second probability distribution function, P_{int}, specified by a different –smaller in principle– parameter set $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots)$. We will denote the external source and its internal representation by $P_{src}(\mathbf{s}|\boldsymbol{\alpha})$ and $P_{int}(\mathbf{s}|\boldsymbol{\beta})$ respectively. The latter aims at capturing the essential features of the former in the most efficient –though in general imperfect– way (see Figure 19A). The external distribution $P_{src}(\mathbf{s}|\boldsymbol{\alpha})$ could therefore be also interpreted as the best representation possible, that typically is not accessible by living systems.

Considering again the example of regulatory networks, the external signals are environmental conditions (temperature, pH,...), that are variable and can only be probabilistically gauged by a cell. The binary vector $\mathbf{s} = (s_1, s_2, \ldots s_N)$ can be thought of as the on/off state of the different N genes in its (Boolean) genetic regulatory network [103, 107, 108]. In this way, $P_{src}(\mathbf{s}|\boldsymbol{\alpha})$ can be interpreted as the probability that the most convenient state aimed by the system to cope with a given environmental condition is $\mathbf{s}$, while $P_{int}(\mathbf{s}|\boldsymbol{\beta})$ is the actual probability for the genetic-network state (attractor) of a given individual –with its limitations– to be $\mathbf{s}$. Without loss of generality, we consider that there is at least one control parameter, say β_1, such that –other parameters being fixed– it determines in which phase the network is operating.

Our thesis is that the capacity of living systems to tune their internal states to efficiently cope with variable external conditions provides them with a strong competitive advantage. Thus, the internal state $P_{int}(\mathbf{s}|\boldsymbol{\beta})$ should resemble as closely as possible the one most in agreement with the environmental signal $P_{src}(\mathbf{s}|\boldsymbol{\alpha})$; in other words, one seeks the distribution that the system should express in order to best respond to the external conditions. Information theory provides us with a robust measure of the "closeness" between the aimed (source) and the actual (internal) probability distribution functions.

Indeed, the Kullback-Leibler (KL) divergence [113], $D(\alpha|\beta)$, quantifies the information loss when the internal state is used to approximate the source (see section 6.2.1). The KL divergence is asymmetric in the two involved probability distributions, it is non-negative, and it vanishes if and only if the two distributions are identical. Minimizing the KL divergence with respect to the internal-state parameters, β, generates the optimal, though in general imperfect, internal state aimed at representing or coping-with a given source (see Fig. 19A).

More generally, in an ever-changing world, the requirement for an individual is not just to reproduce a single source with utmost fidelity but rather to be able to successfully cope with a group of highly diverse sources (see Fig. 19B). A particularly interesting example of this would comprise a community of similar individuals which together strive to establish some kind of a common collective language (see Fig. 19C).

In any of these complex situations, our working hypothesis is that an individual has a larger "fitness" when a characteristic measure, e.g. the mean, of its KL divergences from the set of diverse sources is small, i.e. fit agents are those whose internal states are close to those required by existing external conditions.

We have developed both analytical calculation (see section 6.3) and (see sections 6.4 and 6.5) computational evolutionary models investigating the ideas explained above. The dynamical rules employed in these models are not meant to, necessarily, mimic the actual dynamics of living systems, rather they are efficient ways to optimize fitness. In particular, the result prestented here are also valid in the case of an adaptive dynamics (instead of an evolutionary one) [114] In the evolutionary models a community of M individuals –each one characterized by its own set of internal parameters β– evolves in time through the processes of death, birth, and mutation. Individuals with larger fitness, i.e. with a smaller mean KL divergence from the rest of sources, have a larger probability to produce an offspring, which –apart from small random mutations– inherits its parameters from its ancestor. On the other hand, agents with low fitness are more likely to die and be removed from the community. These evolutionary rules result in the ensemble of agents converging to a steady state distribution, that we characterize in the next sections. We obtain similar results in two families of models, which differ in the way in which the environment is treated. In the first (see section 6.4), the environment is self-generated by a community of co-evolving individuals, while, in the second (see section 6.5), the variable external world is defined *ad hoc*.

As an illustrative example, consider two individual agents A and B –the source for A is B and vice versa– each of them with its own probabilistic gene network. The relative fitnesses of A and B are determined by how well the set of cues (described by the probability

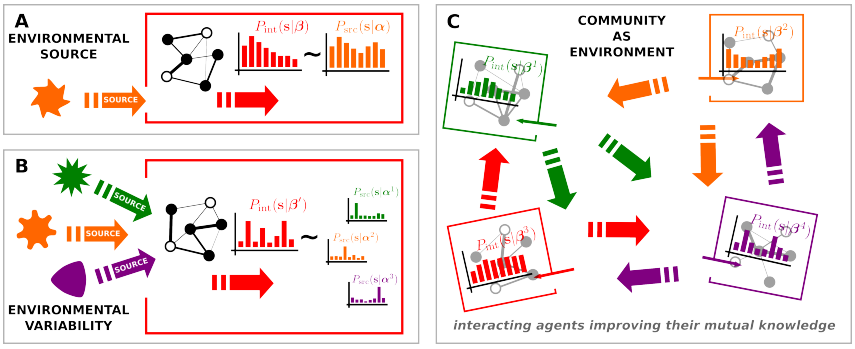


Figure 19: Living systems coping with the environment. Panel A illustrates a *living system* responding to an *environmental source* (e.g. a bacteria responding to some external conditions such as the presence of some nutrients, pH, temperature, ...). A given source, labeled by the set of parameters $\boldsymbol{\alpha}$, can only be probabilistically gauged by the system. $P_{src}(\mathbf{s}|\boldsymbol{\alpha})$ is the most accurate representation that the system can potentially generate in terms of the Boolean variables (or bits) $\mathbf{s}$. However, such a representation might not be accessible to the system by merely changing its internal-state parameters, $\boldsymbol{\beta}$, and the actual internal state, $P_{int}(\mathbf{s}|\boldsymbol{\beta})$, (e.g. the probability of a gene expression pattern) is usually an imperfect proxy for $P_{src}(\mathbf{s}|\boldsymbol{\alpha})$. The optimal choice of parameters $\boldsymbol{\beta}$ –aiming at capturing the most relevant features of the environment– is obtained by minimizing the Kullback-Leibler divergence of $P_{int}(\mathbf{s}|\boldsymbol{\beta})$ from $P_{src}(\mathbf{s}|\boldsymbol{\alpha})$ (see section 6.2.1). In genetic networks, changing internal parameters is equivalent to changing the interactions between the different (Boolean) variables (nodes of the networks in the figure). Panel B shows a more complex scenario, where the system has to cope with multiple and diverse sources. The internal state has to be able to accommodate each of them. In panel C, the environment is not imposed *ad hoc* but instead, it is composed of other individuals, and every agent needs to cope with ("understand") the states of the others. Each agent evolves similarly to the others in the community, trying to exhibit the same kind of state, generating in this way a self-organized environment.

distribution P_{src}) of one organism is captured by the other with minimum information loss, and vice versa (for utter simplicity, we could assume that the distributions associated with A and B correspond to equilibrium distributions of an Ising model [115] at similar inverse temperatures β_A and β_B). If $\beta_A = \beta_B$, the two distributions would be identical and the KL divergence would vanish. However, this is not a stable solution. Indeed, if the two parameters are not identical but close, the difference between their respective KL divergences from each to the other is (see section 6.3):

$$D_{KL}(\beta_A + \delta\beta|\beta_A) - D_{KL}(\beta_A|\beta_A + \delta\beta) \simeq \frac{1}{6}\nabla\chi(\beta_A)\delta\beta^3 \,, \quad (91)$$

where χ is the generalized susceptibility also known as "Fisher information" (see section 6.2.2). This implies that the individual whose parameters correspond to the state with larger χ has a smaller KL divergence and is thus fitter. But it is well-known that χ peaks at the critical point, and thus our key finding is that, for a family of individuals with similar parameters, the fittest possible agent sits exactly at criticality, and it is best able to encapsulate a wide variety of distributions. As we illustrate in what follows with a number of examples, the optimal encoding parameters of stable solutions lie always around the peak of the generalized susceptibility χ which is the region of maximal variability, where different complex sources can be best accounted for through small parameter changes.

6.2.1 *Information theory and Kullback-Leibler divergence*

Given two probability distributions $P(\mathbf{s})$ and $Q(\mathbf{s})$ the Kullback-Leibler (KL) divergence of $Q(\mathbf{s})$ from $P(\mathbf{s})$ is defined as

$$D(P(\cdot)|Q(\cdot)) := \sum_{\mathbf{s}} P(\mathbf{s}) \log \frac{P(\mathbf{s})}{Q(\mathbf{s})} \,, \quad (92)$$

and quantifies the loss of information obtained when $Q(\mathbf{s})$ is used to approximate $P(\mathbf{s})$ [116, 113]. The KL divergence is non-negative and vanishes if and only if the two distributions are equal. It is important to stress that the KL divergence is not symmetric and therefore is not a properly-defined distance.

The KL divergence can be better understood when related to maximum likelihood principle via the Sanov's theorem [117, 113, 118]. Consider a long sequence of empirical data consisting of T independent measurements. Let $C(\mathbf{s})$ be the number of times a certain event $\mathbf{s}$ is repeated in the sequence. Suppose that the actual, though unknown, empirical distribution of those events is $P(\mathbf{s})$. In the large T limit, the frequencies $C(\mathbf{s})/T$ converge to $P(\mathbf{s})$, by the Glivenko-

Cantelli theorem [113]. Let then $Q(\mathbf{s})$ be probability distribution that defines a model The (multinomial) likelihood is defined as

$$\mathcal{L} = \frac{T!}{\prod_{\mathbf{s}} C(\mathbf{s})!} \prod_{\mathbf{s}} Q(\mathbf{s})^{C(\mathbf{s})} , \tag{93}$$

which is nothing but the probability that the model $Q(\mathbf{s})$ generates the T observations (i.e. $C(\mathbf{s})$). The previous equation can be rewritten as

$$\mathcal{L} = T! \, \exp \sum_{\mathbf{s}} \Big(C(\mathbf{s}) \log \big(Q(\mathbf{s}) \big) - \log \big(C(\mathbf{s})! \big) \Big) , \tag{94}$$

which, in the large T limit, using $T! \to T^T$ and $C(\mathbf{s}) \to TP(\mathbf{s})$ becomes

$$\mathcal{L} \sim \exp\Big(-T D\big(P(\cdot)|Q(\cdot)\big) \Big) , \tag{95}$$

up to leading order. Therefore, maximizing the likelihood of a trial probability distribution function Q is equivalent to minimizing its KL divergence with respect to the original one, P. This result is also known as Sanov's theorem [117] in the context of large deviations theory.

6.2.2 *Fisher Information*

Given a probability distribution $P(\mathbf{s}|\boldsymbol{\gamma})$ –where γ is a set of parameters – the Fisher Information is defined as

$$\chi^{\mu\nu}(\boldsymbol{\gamma}) := \left\langle \frac{\partial \log P(\cdot|\boldsymbol{\gamma})}{\partial \gamma_\mu} \frac{\partial \log P(\cdot|\boldsymbol{\gamma})}{\partial \gamma_\nu} \right\rangle_{\boldsymbol{\gamma}} , \tag{96}$$

where μ and ν are parameter labels and the average $\langle \cdot \rangle_{\boldsymbol{\gamma}}$ is performed with respect to $P(\cdot|\boldsymbol{\gamma})$. It measures the *amount of information* encoded in the states $\mathbf{s}$ about the parameters $\boldsymbol{\gamma}$ [113]. This follows from the Cramér-Rao inequality, which states that the error made when we estimate $\boldsymbol{\gamma}$ from one state $\mathbf{s}$ is, on average, greater (or at least equal) than the inverse of the Fisher information [113]. In particular, if χ happens to diverge at some point, it is possible to specify the associated parameters with maximal precision [99].

A generic probability distribution can be rewritten to parallel the standard notation in statistical physics

$$P(\mathbf{s}|\boldsymbol{\gamma}) = \exp\big(-H(\mathbf{s}|\boldsymbol{\gamma}) \big) / Z(\boldsymbol{\gamma}) , \tag{97}$$

where the factor $Z(\boldsymbol{\gamma})$ is fixed through normalization. The function H can be generically written as

$$H(\mathbf{s}|\boldsymbol{\gamma}) = \sum_{\mu} \gamma_\mu \phi^\mu(\mathbf{s}) , \tag{98}$$

where $\phi^\mu(\mathbf{s})$ are suitable functions ("observables") of the variables $\mathbf{s}$.

With the parametrization used in equation 97, the Fisher information is the *generalized susceptibility* in statistical mechanics terminology and measures the response of the system to parameter variations:

$$\chi^{\mu\nu}(\boldsymbol{\gamma}) = -\frac{\partial\langle\phi^\mu\rangle_\gamma}{\partial\gamma_\nu} = \langle\phi^\mu\phi^\nu\rangle_\gamma - \langle\phi^\mu\rangle_\gamma\langle\phi^\nu\rangle_\gamma \,, \tag{99}$$

and is well-known to peak at critical points [115].

6.3 ANALYTICAL RESULTS

The KL divergence between the distributions characterized by generic parameter sets $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively, $D(\boldsymbol{\alpha}|\boldsymbol{\beta}) = D(P_{src}(\cdot|\boldsymbol{\alpha})|P_{int}(\cdot|\boldsymbol{\beta}))$, can be easily written by using the generic parametrization of equations 97 and 98:

$$D(\boldsymbol{\alpha}|\boldsymbol{\beta}) = \sum_{\mu=1}^{I} \beta_\mu\langle\phi^\mu_{int}\rangle_{\boldsymbol{\alpha}} + \log Z_{int}(\boldsymbol{\beta}) - S_{src}(\boldsymbol{\alpha}) \,, \tag{100}$$

with

$$\langle\phi^\mu_{int}\rangle_{\boldsymbol{\alpha}} := \sum_{\mathbf{s}} \phi^\mu_{int}(\mathbf{s})P_{src}(\mathbf{s}|\boldsymbol{\alpha}) \tag{101}$$

$$S_{src}(\boldsymbol{\alpha}) := -\sum_{\mathbf{s}} P_{src}(\mathbf{s}|\boldsymbol{\alpha})\log\Big(P_{src}(\mathbf{s}|\boldsymbol{\alpha})\Big) \,, \tag{102}$$

where the last expression is the entropy of the distribution P_{src}.

To proceed further, we have to define the optimal choice of parameter $\boldsymbol{\beta}$ given the distribution of external parameters $\rho_{src}(\boldsymbol{\alpha})$.

We define the optimal value of $\boldsymbol{\beta}$ as the one having the lowest average KL divergence over the parameters $\boldsymbol{\alpha}$. Formally, it is written as

$$\boldsymbol{\beta}_a^{opt} = \arg\min_{\boldsymbol{\beta}}\Big[\int d\boldsymbol{\alpha} D(\boldsymbol{\alpha}|\boldsymbol{\beta})\Big)\rho_{src}(\boldsymbol{\alpha})\Big] \tag{103}$$

and the system adopts the single value of $\boldsymbol{\beta}$, $\boldsymbol{\beta}_a^{opt}$, that better describes *on average* the varying environment.

Here we analyze the model defined via equation 103, the optimal choice corresponds to the value of the internal parameter $\boldsymbol{\beta}$ which minimizes the average KL divergence to the sources $\boldsymbol{\alpha}$, defined as

$$d(\rho_{src}|\boldsymbol{\beta}) := \int d\boldsymbol{\alpha}\rho_{src}(\boldsymbol{\alpha})D(\boldsymbol{\alpha}|\boldsymbol{\beta}) \,. \tag{104}$$

Using equation 100, into this, we obtain an equation for the stationary points

$$0 = \frac{\partial}{\partial\beta_\mu} d(\rho_{src}|\boldsymbol{\beta}) = -\langle\phi^\mu_{int}\rangle_{\boldsymbol{\beta}} + \int d\boldsymbol{\alpha}\ \rho_{src}(\boldsymbol{\alpha})\ \langle\phi^\mu_{int}\rangle_{\boldsymbol{\alpha}} \,. \tag{105}$$

This equation can be interpreted in an alternative way; introducing the "averaged environment"

$$\bar{P}_{src}(\mathbf{s}|\rho_{src}) := \int d\boldsymbol{\alpha} \rho_{src}(\boldsymbol{\alpha}) P_{src}(\mathbf{s}|\boldsymbol{\alpha}) \,, \tag{106}$$

the KL divergence respect to $P_{int}(\mathbf{s}|\boldsymbol{\beta})$ is

$$D(\bar{P}_{src}(\cdot|\rho_{src})|P_{int}(\cdot|\boldsymbol{\beta})) = d(\rho_{src}|\boldsymbol{\beta}) - \int d\boldsymbol{\alpha} \rho_{src}(\boldsymbol{\alpha}) D(P_{src}(\cdot|\boldsymbol{\alpha})|\bar{P}_{src}(\cdot|\rho_{src})) \,. \tag{107}$$

Since the last term on the right hand side does not depend on $\boldsymbol{\beta}$, the minimization of the KL divergence between the "averaged environment" and the internal mapping $P_{src}(\mathbf{s}|\boldsymbol{\beta})$ leads to the same result as the minimization of $d(\rho_{src}|\boldsymbol{\beta})$ given by eq. 104. In both cases, the Hessian matrix turns out to be strictly positive, and therefore the stationary points are local minima.

In the particular case in which the internal and the source probabilities $P_{ind}(\mathbf{s}|\boldsymbol{\alpha})$ and $P_{src}(\mathbf{s}|\boldsymbol{\beta})$ have the same structure, i.e. $\phi^{\mu}_{src} = \phi^{\mu}_{int} = \phi^{\mu}$, the optimal internal parameters $\boldsymbol{\beta}$ is given by the solution of eq. (105), which can be simply written as

$$\left\langle \phi^{\mu} \right\rangle_{\boldsymbol{\beta}} = \int d\boldsymbol{\alpha} \, \rho_{src}(\boldsymbol{\alpha}) \left\langle \phi^{\mu} \right\rangle_{\boldsymbol{\alpha}} \,. \tag{108}$$

We then study how the distance to the critical point is modified by an internal representation. We consider a narrow distribution of $\boldsymbol{\alpha}$ values, characterized by a given average value, $\bar{\boldsymbol{\alpha}}$, at some distance to criticality, and investigate how do the distance to criticality changes when the optimal internal representation is build.

For this purpose, we analyze the local behavior of the internal parameters, taking a distribution $\rho_{src}(\boldsymbol{\alpha})$ which is significantly different from zero only in a small region U. We can therefore expand both of the left and the right hand sides of eq. 108 around the mean value $\bar{\boldsymbol{\alpha}} := \int_U d\boldsymbol{\alpha} \, \rho_{src}(\boldsymbol{\alpha}) \boldsymbol{\alpha}$, and then we obtain

$$(\beta_{\mu} - \bar{\alpha}_{\mu}) = \left(\chi^{-1}(\bar{\boldsymbol{\alpha}})\right)^{\mu\nu} \left(\frac{1}{2} \Big| \frac{\partial}{\partial \alpha_{\nu}} \chi^{\gamma\delta}(\boldsymbol{\alpha}) \Big|_{\bar{\boldsymbol{\alpha}}} \right) \int_U d\boldsymbol{\alpha}' \rho_{src}(\boldsymbol{\alpha}') (\alpha'_{\gamma} - \bar{\alpha}_{\gamma})(\alpha'_{\delta} - \bar{\alpha}_{\delta}) \,, \tag{109}$$

where we are summing over repeated indices and $\chi^{\mu\nu}(\boldsymbol{\alpha})$ is defined in section 6.2.2. This equation quantifies the deviation of internal parameters $\boldsymbol{\beta}$ respect to $\bar{\boldsymbol{\alpha}}$. To understand its relation with the critical point, we rewrite eq. 109 as

$$(\boldsymbol{\beta} - \bar{\boldsymbol{\alpha}}) = \chi^{-1}(\bar{\boldsymbol{\alpha}}) \cdot \nabla \Omega(\bar{\boldsymbol{\alpha}}), \tag{110}$$

where we have defined the scalar field $\Omega(\boldsymbol{\alpha})$ as

$$\Omega(\boldsymbol{\alpha}) := \frac{1}{2}\int_U d\boldsymbol{\alpha}'\, \rho_{src}(\boldsymbol{\alpha}')(\boldsymbol{\alpha}'-\bar{\boldsymbol{\alpha}})^T\cdot\chi(\boldsymbol{\alpha})\cdot(\boldsymbol{\alpha}'-\bar{\boldsymbol{\alpha}})\,. \tag{111}$$

As $\chi(\boldsymbol{\alpha})$ is a symmetric positive-definite matrix, $\Omega(\boldsymbol{\alpha})$ is a positive quantity. If we introduce a base of eigenvectors of $\chi(\boldsymbol{\alpha})$, $\mathbf{v}^\gamma(\boldsymbol{\alpha})$ with eigenvalues $\lambda^\gamma(\boldsymbol{\alpha})$, we get

$$\Omega(\boldsymbol{\alpha}) = \frac{1}{2}\lambda^\gamma(\boldsymbol{\alpha})\int_U d\boldsymbol{\alpha}'\, \rho_{src}(\boldsymbol{\alpha}')\left[\mathbf{v}^\gamma(\boldsymbol{\alpha})^T\cdot(\boldsymbol{\alpha}'-\bar{\boldsymbol{\alpha}})\right]^2\,. \tag{112}$$

At the critical point, at least one $\lambda^\gamma(\boldsymbol{\alpha})$ diverges (in the thermodynamic limit), so $\Omega(\boldsymbol{\alpha})$ has a maximum at the critical point. Note that $\mathbf{v}^\gamma(\boldsymbol{\alpha})$ cannot vanish because it is an unitary vector. Therefore, the gradient of Ω points to the critical point (at least if $\bar{\boldsymbol{\alpha}}$ is not too far from it in a such a way that there are not other local maxima).

Finally, we project both sides of eq. 110 over the gradient of $\Omega(\bar{\boldsymbol{\alpha}})$, obtaining

$$(\nabla\Omega(\bar{\boldsymbol{\alpha}}))^T\cdot(\boldsymbol{\beta}-\bar{\boldsymbol{\alpha}}) = (\nabla\Omega(\bar{\boldsymbol{\alpha}}))^T\cdot\chi^{-1}(\bar{\boldsymbol{\alpha}})\cdot\nabla\Omega(\bar{\boldsymbol{\alpha}})\,. \tag{113}$$

As χ is positive-definite, also its inverse χ^{-1} is a positive-definite matrix, and the projection of $(\boldsymbol{\beta}-\bar{\boldsymbol{\alpha}})$ on the gradient of $\Omega(\bar{\boldsymbol{\alpha}})$ (which points to the critical point) is also positive.

Consequently, the internal parameters α that are closer to the critical point than $\bar{\boldsymbol{\alpha}}$ have larger fitness, indicating that there is an overall drift towards parameter regions with larger Fisher information, i.e. toward criticality. Equation 91 is a simple example of this general case, where just two individuals are considered instead of a continuous distribution.

6.3.1 *Example: Ising mean field*

We analyze numerically a simple case, inspired by the archetypical (mean-field) Ising model, in which

$$H_{src}(\mathbf{s}|\boldsymbol{\alpha}) = -\frac{N}{2}\alpha_1\left(\sum_i\frac{s_i}{N}\right)^2 - \frac{N}{4!}\alpha_2\left(\sum_i\frac{s_i}{N}\right)^4 \tag{114}$$

and

$$H_{int}(\mathbf{s}|\boldsymbol{\beta}) = -\frac{N}{2}\beta_1\left(\sum_i\frac{s_i}{N}\right)^2 = -\frac{1}{N}\beta_1\sum_{i,j>i}^{N}s_is_j + \text{constant}. \tag{115}$$

Numerical factors $(-N/2)$ and $(-N/4!)$ have been introduced for convenience. The last form of the H_{int} visually relates this internal representation with the classic problem of Boltzmann learning [119].

We take a specific distribution $\rho_{src}(\boldsymbol{\alpha})$ –in this case a uniform distribution in the range $[-10, 10]$– and we compute the mean distance

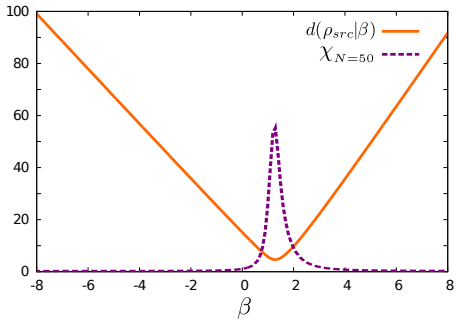


Figure 20: *Annealed choice.* The orange line is the averaged KL divergence $d(\rho_{src}|\beta)$ of the internal parameters β, defined in eq. 104. The dashed line is proportional to the susceptibility of the internal model. The minimum of the distance is located close to the critical point (defined by the peak of the generalized susceptibility). In this case both the source and the internal representations are modeled by equation 115 with $N = 50$, and the source parameters are uniformly distributed in the range $[-10, 10]$.

between a distribution parametrized by an α (drawn from ρ_{src}) and an internal parameter β. Fig. 20 show that the parameter β that minimizes this distance is close to the one that maximize the Fisher information, i.e. the critical point.

,

6.4 CO-EVOLUTIONARY MODEL

The environment perceived by each individual consists of the other $M - 1$ systems in the community, which it aims at "understanding" and coping with. In the simplest computational implementation of this idea, a pair of individual agents is randomly selected from the community at each time step and each of these two individuals constitutes the environmental source for the other. Given that the KL divergence is not symmetric (see section 6.2.1), one of the two agents has a larger fitness and thus a greater probability of generating progeny, while the less fit system is more likely to die. This corresponds to a fitness function of agent i which is a decreasing function of the KL divergence from the other. In this case, as illustrated in Fig. 21, the co-evolution of $M = 100$ agents –which on their turn are sources– leads to a very robust evolutionarily-stable steady-state distribution. Indeed, the different left panels show that for three substantially different initial parameter distributions (very broad, and localized in the ordered and in the disordered phases, respectively) the community co-evolves in time to a unique localized steady state distribution, which turns out to be peaked at the critical point (i.e. where the Fisher information peaks, see Fig. 21 right panel). This conclusion is robust against model details and computational implementations: the solution peaked at criticality is an evolutionary stable attractor

of the dynamics. The same conclusions hold for an analogous "Co-adaptive model" in which the systems adapt rather than dying and replicating [114].

In this model the k-th agent of the community is described by a probability distribution $P_{int}(\mathbf{s}|\boldsymbol{\beta}^k) \propto \exp\{-H_{int}(\mathbf{s}|\boldsymbol{\beta}^k)\}$, depending on parameters $\boldsymbol{\beta}^k$.

We consider as initial state an ensemble of M individuals whose coupling parameters are extracted from an arbitrary distribution, $p(\boldsymbol{\beta}, t = 0)$. The evolutionary dynamics proceeds as follows:

1. At each time step, two individuals, i and j, are randomly selected.
2. Their relative fitnesses $f_i^{(j)}$ and $f_j^{(i)}$ are defined as

$$\begin{aligned} f_i^{(j)} &= D(\boldsymbol{\beta}^i|\boldsymbol{\beta}^j) = \sum_{\mathbf{s}} P_{int}(\mathbf{s}|\boldsymbol{\beta}^i) \log \frac{P_{int}(\mathbf{s}|\boldsymbol{\beta}^i)}{P_{int}(\mathbf{s}|\boldsymbol{\beta}^j)} \\ f_j^{(i)} &= D(\boldsymbol{\beta}^j|\boldsymbol{\beta}^i) = \sum_{\mathbf{s}} P_{int}(\mathbf{s}|\boldsymbol{\beta}^j) \log \frac{P_{int}(\mathbf{s}|\boldsymbol{\beta}^j)}{P_{int}(\mathbf{s}|\boldsymbol{\beta}^i)} \,, \end{aligned} \tag{116}$$

   i.e. the respective KL divergences. As the KL divergence is not symmetric, $f_i^{(j)} \neq f_j^{(i)}$ unless $\boldsymbol{\beta}^i = \boldsymbol{\beta}^j$.
3. An offspring of one of the two individuals –selected with probability proportional to its relative fitness– is added, while the other individual is removed from the community.
4. Offspring mutate with a probability ν, modifying its parameters from $\boldsymbol{\beta}$ to $\boldsymbol{\beta} + \boldsymbol{\xi}$, where $\boldsymbol{\xi}$ is randomly drawn from a multivariate Gaussian distribution with zero mean and standard deviation σ.
5. Time is updated to $t \to t + 1/M$.
6. Another pair of individuals i and j is picked, and the process is iterated.

To compute the stationary distribution of parameters $p(\boldsymbol{\beta}) \equiv p(\boldsymbol{\beta}, t \to \infty)$, we iterate T_i time steps and then perform measurements during $T_f - T_i$ steps. Results are averaged over R realizations of the evolutionary process.

We now present some specific realizations of these general co-evolutionary rules. The "internal" probability distributions of a single individual are considered of the following form

$$H_{int}(\mathbf{s}|\boldsymbol{\beta}) = -\frac{N}{2}\beta_1 \left(\sum_i \frac{s_i}{N}\right)^2 - N\beta_2 \left(\sum_i \frac{s_i}{N}\right) \tag{117}$$

with a linear and a quadratic coupling. We will also discuss the simpler, particular case in which β_2 vanishes (i.e. only the quadratic coupling is present), as well as the quadratic-quartic case:

$$H_{int}(\mathbf{s}|\boldsymbol{\beta}) = -\frac{N}{2}\beta_1\left(\sum_i \frac{s_i}{N}\right)^2 - \frac{N}{4!}\beta_2\left(\sum_i \frac{s_i}{N}\right)^4 \tag{118}$$

Parameter	**Value**
N	100
M	100
ν	0.1
σ	0.1
σ_1, σ_2	$1\sigma, 0.1\sigma$
T_i	10^4
$T_f - T_i$	10^5
Init. Distribution	$N(-3, 0.25) \cdot N(-0.25, 0.05)$
	$N(3, 0.25) \cdot N(0.25, 0.05)$
	$U([-4, 4] \times [-0.8, 0.8])$
R	100 (10000 for transients)

Table 1: Parameter values in the simulation of the Co-evolutionary Model in Figure 21. See also Table 2.

6.4.0.1 *Results*

As for the Evolutionary Model, we first study the stationarity of the final distribution of parameters $\boldsymbol{\beta}$ in the community and its dependence on the number of spins, N, number of individuals, M, the mutation probability, ν, and the deviation of the mutations, σ.

Figure 21 shows the distribution of individual parameters, $p(\boldsymbol{\beta})$ for the linear-quadratic case, eq. 117, for three different initial distributions and different transient periods, together with the final stationary distribution which is the same for all of them. Numerical parameters are summarized in Table 1.

The main result is very robust under reparametrization: results for the case of a community with individuals whose probability distribution is characterized by the linear-quadratic case, eq. 117 are summarized in Figs. 22 and 23, whereas Figs. 24 and 25 correspond to the one-parameter case (eq. 115) and the quadratic-quartic case (eq. 118), respectively.

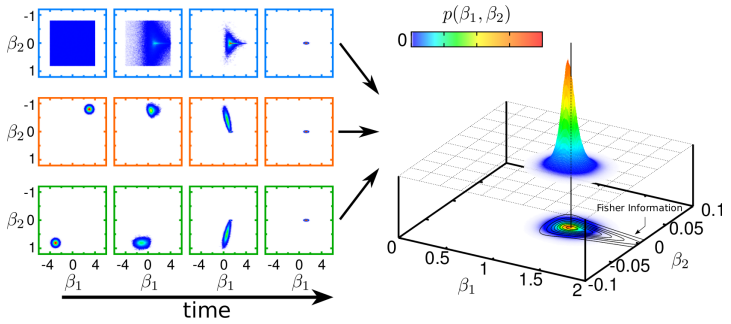


Figure 21: **Co-evolutionary model leads self-consistently to criticality:** A community of M living systems (or agents) evolves according to a genetic algorithm dynamics [120]. Each agent i ($i = 1, ..., M$) is characterized by a 2-parameter (β_1^i, β_2^i) internal state distribution $P_{int}(\mathbf{s}|\beta_1^i, \beta_2^i)$ and the rest of the community acts as the external environment it has to cope with, i.e. the agents try to "understand" each other. At each simulation step, two individuals are randomly chosen and their respective relative fitnesses are computed in terms of the Kullback-Leibler divergence from each other's internal state probability distribution. One of the two agents is removed from the community with a probability that is smaller for the fitter agent; the winner produces an offspring which (except for small variations/mutations) inherits its parameters. These co-evolutionary rules drive the community very close to a unique localized steady state. As shown in the right panel, this is localized precisely at the critical point, i.e. where the generalized susceptibility or Fisher information of the internal state distribution exhibits a sharp peak (as shown by the contour plots and heat maps). The internal state distributions are parameterized as $P_{int}(\mathbf{s}|\beta_1, \beta_2) \propto \exp\{\beta_1 \frac{N}{2}(\sum_{k=1}^{N} \frac{s_k}{N})^2 + \beta_2 \sum_{k=1}^{N} s_k\}$ representing a global (all-to-all) coupling of the internal nodes (see MM). Much more complex probability distributions in which all units are not coupled to all other units –i.e. more complex networked topologies– are discussed in the SI Appendix, sec. S4.

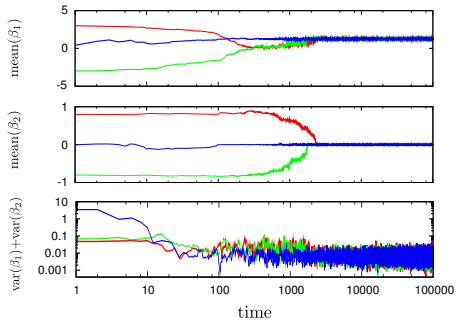


Figure 22: *Time evolution in the Co-evolutionary Model* (in the linear-quadratic case, i.e. eq. 117). Plot of the mean values $\langle \beta_{1,2} \rangle := \frac{1}{M} \sum_{k=1}^{M} \beta_{1,2}^{k}$ (top two panels) of the community, and parameter variance (bottom panel) for three different realizations started with different initial conditions $p(\beta, t = 0)$: Gaussian distributions with two different averages and variances, $N(-3, 0.25)N(-0.8, 0.05)$ (red line) and $N(3, 0.25)N(0.8, 0.05)$ (green line), and uniform distribution in the range $[-4, 4] \times [-0.8, 0.8]$ (blue line).

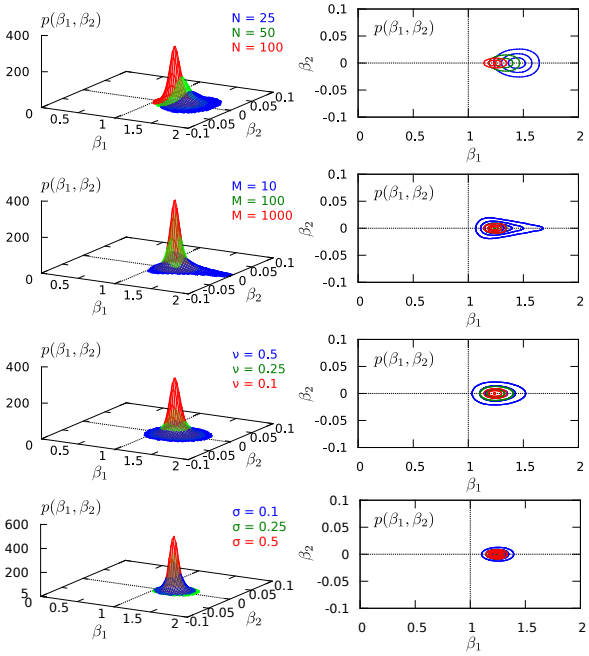


Figure 23: *Parameter dependence in the Co-evolutionary Model:* stationary distribution $p(\beta_1, \beta_2)$ as a function of parameters (for the linear-quadratic case, eq. 117). Different colored lines in each plot correspond to different values of N, community sizes M, mutation parameters ν, and σ. For larger communities sizes the stationary distribution becomes sharper. Parameter values are listed in Table 1 (unless otherwise specified).

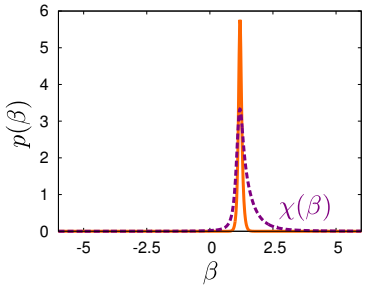


Figure 24: *Stationary distribution* $p(\beta)$ *in the Co-evolutionary Model, eq. 115:* we compare the stationary distribution (orange line) with the generalized susceptibility (purple dashed line). As in the example in the main text, the individual parameters converge to the neighborhood of the peak of the Fisher information. Parameters are set to $N = 100$, $M = 100$, $\nu = 0.1$, $\sigma = 0.1$.

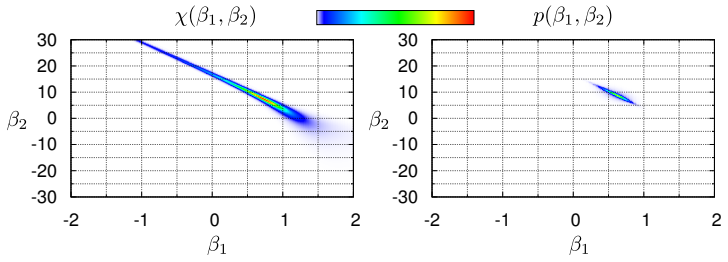


Figure 25: *Stationary distribution* $p(\beta_1, \beta_2)$ *in the Co-evolutionary Model, eq. 118:* as above, we compare the stationary distribution (right panel) with the generalized susceptibility (left panel). Again, the community evolves toward the global maximum of the Fisher information. Parameter values: $N = 100$, $M = 100$, $\nu = 0.1$, and $\sigma_1 = \sigma_2 = 0.1$.

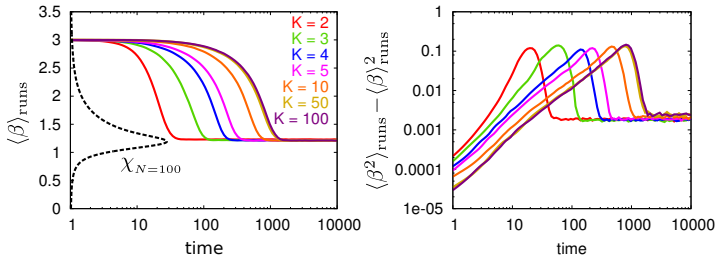


Figure 26: *Time evolution in the Co-evolutionary Model with K-body interactions* (with parametrization of eq. 115). Solid lines represent the time evolution of mean values $\langle\beta\rangle := \int d\beta\ p(\beta,t)\beta$. The relaxation to the stationary state depends on the effective number of individuals K with which each single agent interacts. The larger the value of K the larger the relaxation time. The community evolves very close to the maximum of the generalized susceptibility (plotted with dashed lines). The initial condition is $\beta = 3$ for all of the individuals and parameters are $N = 100$, $M = 100$, $\nu = 1$ and $\sigma = 0.1$.

6.4.1 *Co-evolutionary Model with K-body interactions*

In the previous section we considered the case of pairs of individuals considered at each time step. In this section we study a variant of the co-evolutionary model in which K individuals $(i_1, ..., i_K)$ are randomly picked at each time step. The probability of each individual to die is proportional to its (normalized) average KL divergence to the remaining ones, i.e.

$$P_{\rm kill}(i_k) = \frac{\sum_{l=1}^{K} D(\boldsymbol{\beta}^{i_l}|\boldsymbol{\beta}^{i_k})}{\sum_{m=1}^{K}\sum_{l=1}^{K} D(\boldsymbol{\beta}^{i_l}|\boldsymbol{\beta}^{i_m})} \,. \tag{119}$$

When one individual dies it is replaced by a copy of one of the remaining $K-1$ individuals (and mutations are introduced with probability ν).

We have implemented the simulation with parametrization of eq. 115. Results are summarized in Fig. 26 where we plot the mean value of the parameter over the entire community and 10^3 realizations of the same initial condition. It can be seen that the time to reach stationarity increases with K. When K increases, the drift which moves the system towards the criticality is lower. This is related to the fact that, by averaging over more and more individuals, the source effectively becomes more and more homogeneous.

6.4.2 *Co-evolutionary Model with complex internal networked topologies*

We now consider a different variant of the model in which the internal probability distribution of each individual/agent is not mean-field, in the sense that every s_i variable is coupled to all others. Instead, possible interactions are encoded in a network, such that each s_i interacts only with other s_j directly connected to it, i.e. for which the *adjacency matrix*, element $a_{ij} \neq 0$.

The evolutionary dynamics is as above, with the only difference that now the structure of the probability characterizing each individual is as follows:

- Given N spin variables, we generate a fixed adjacency matrix of interactions $\hat{a}$. The probability to find a certain configuration $\mathbf{s}$ in the k-th individual is $P^{\hat{a}}_{\text{int}}(\mathbf{s}|\beta^k) \propto \exp\{-H^{\hat{a}}_{\text{int}}(\mathbf{s}|\beta^k)\}$ with

$$H^{\hat{a}}_{\text{int}}(\mathbf{s}|\beta^k) = -\beta^k \frac{1}{N} \sum_{i,j>i}^{N} a_{ij} s_i s_j. \tag{120}$$

- The system is iterated as in the Co-evolutionary Model, leaving $\hat{a}$ fixed in time and identical for all individuals.

As the structure of $\hat{a}$ is an arbitrary one, the calculation of the distances between distributions needs to be explicitly computed by summing over the 2^N possible states (which severely limits the maximum size in computer simulations; $N \sim 20$).

Results for the stationary distribution of parameters $p(\beta)$ and different types of network architectures, together with the corresponding curves of generalized susceptibilities computed as

$$\chi^{\hat{a}}(\beta) = \left\langle \left(\frac{1}{N} \sum_{i,j>i}^{N} a_{ij} s_i s_j \right)^2 \right\rangle_{P^{\hat{a}}(\mathbf{s}|\beta)} - \left\langle \frac{1}{N} \sum_{i,j>i}^{N} a_{ij} s_i s_j \right\rangle^2_{P^{\hat{a}}(\mathbf{s}|\beta)} \tag{121}$$

are shown in Fig. 27. In all cases, the main result of this chapter holds: the resulting internal parameters distributions peak around the critical point.

6.5 EVOLUTIONARY MODEL

We model a community of individuals receiving external stimuli from an outer and heterogeneous environment. We describe every source of the environment by a generic distribution, as parametrized by equation 97 with

$$H_{\text{src}}(\mathbf{s}|\boldsymbol{\alpha}) = \sum_{\mu}^{E} \alpha_\mu \phi^\mu_{\text{src}}(\mathbf{s}), \tag{122}$$

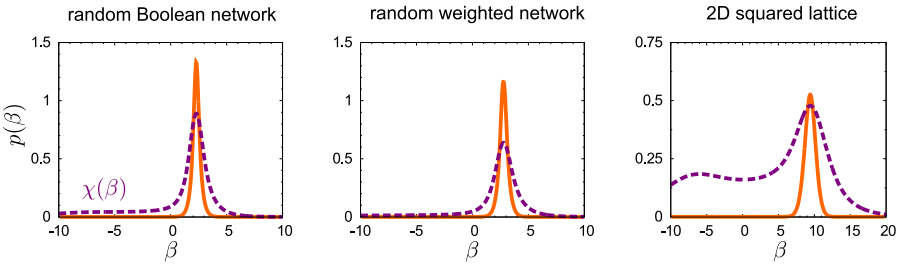


Figure 27: Stationary parameter distributions for three different network structures $\hat{a}$ of N nodes. Three cases are studied: **(left) Random Boolean network**: connections are not weighted, i.e. $a_{ij} = a_{ji} = \{0, 1\}$, with mean connectivity $N/2$, **(center) Random weighted network**: in this case, $a_{ij} = a_{ji} = \eta$, where η is a random number between 0 and 1**(right) Regular 2D lattice** with periodic bounding conditions. In all cases, parameters have been set to $\nu = 0.1$, $\sigma = 0.5$, $M = 100$. In the first two examples, $N = 20$, while $N = 25$ for last one. In dashed line, the generalized susceptibility has been plotted and re-scaled for visual comparison.

where the parameters $\boldsymbol{\alpha}$ are drawn from the distribution $\rho_{src}(\boldsymbol{\alpha})$.

In the community, each agent has a representation of the observed source (constructed as explained in previous sections). The internal state of the k-th agent is modeled with an internal system described by equation 97, with

$$H_{int}(\mathbf{s}|\boldsymbol{\beta}^k) = \sum_{\mu}^{I} \beta_\mu^k \phi_{int}^\mu(\mathbf{s}), \tag{123}$$

We consider the following dynamics:

1. We start with M individuals each one characterized by its parameters drawn from some arbitrary (broad) distribution. $p(\boldsymbol{\beta}, t = 0)$.

2. At every time step, we generate S external sources, $\{\boldsymbol{\alpha}^u\}_{u=1,\dots,S}$, from the distribution $\rho_{src}(\boldsymbol{\alpha})$.

3. We compute the average KL divergence of every individual's internal representation to the external sources

$$d(\{\boldsymbol{\alpha}^u\}|\boldsymbol{\beta}^k) := \frac{1}{S}\sum_{u=1}^{S}\sum_{\mathbf{s}} P_{src}(\mathbf{s}|\boldsymbol{\alpha}^u)\log\frac{P_{src}(\mathbf{s}|\boldsymbol{\alpha}^u)}{P_{int}(\mathbf{s}|\boldsymbol{\beta}^k)}. \tag{124}$$

4. One of the individuals of the community is removed with a probability proportional to its average KL divergence

$$P_{kill}(k) = \frac{d(\{\boldsymbol{\alpha}^u\}|\boldsymbol{\beta}^k)}{\sum_l d(\{\boldsymbol{\alpha}^u\}|\boldsymbol{\beta}^l)} \tag{125}$$

and it is replaced by a copy of another individual (offspring), which is picked randomly with uniform probability.

5. The offspring inherits its parameter set from its parent or, instead, mutates with a probability ν, altering the original parameter set, $\boldsymbol{\beta} \to \boldsymbol{\beta} + \boldsymbol{\xi}$, where $\boldsymbol{\xi}$ is a random Gaussian number of zero mean and deviation σ.

6. Time is incremented as $t \to t + 1/M$.

7. Another set of parameters $\{\boldsymbol{\alpha}^u\}_{u=1,\dots,S}$ is generated from $\rho_{src}(\boldsymbol{\alpha})$, and the process is iterated.

We are interested in measuring the stationary distribution of the individual parameters, $p(\boldsymbol{\beta}) \equiv p(\boldsymbol{\beta}, t \to \infty)$ when $t \to \infty$ (we start measuring at some time T_i and stop at time T_f), for which the distribution is averaged over R independent realizations of the initial distribution $p(\boldsymbol{\beta}, t = 0)$.

We have simulated the simple case in which both the external sources and internal representations correspond to the simple choice given in equation 114 with $\alpha_2 = 0$ and equation 115 respectively. Similar results can be obtained for other parametrization of sources and internal representations, for instance by considering equation 114 with non-vanishing α_2.

Numerical computation of the Kullback-Leibler divergence

We explore the dependence of the stationary distribution $p(\beta)$ on parameter values (see Table 2).

Parameter	**Value**
N	100
M	100
S	10
ν	0.1
σ	0.1
$\rho_{src}(\alpha_1)$	$\mathcal{U}([-10, 10])$
(not used in Fig. 3 of main text)	
T_i	10^4
$T_f - T_i$	10^5
R	100

Table 2: *Parameters of the simulation of the Evolutionary Model in Fig. 3 of the main text and Fig. 29:* N is the number of spins composing each of the individuals, M is the community size, S is the number of stimuli received in every interaction with the environment, ν is the mutation probability, σ is the deviation of the mutated offspring, T_i and T_f are the initial and final time steps used for the measure and R is the number of independent realizations.

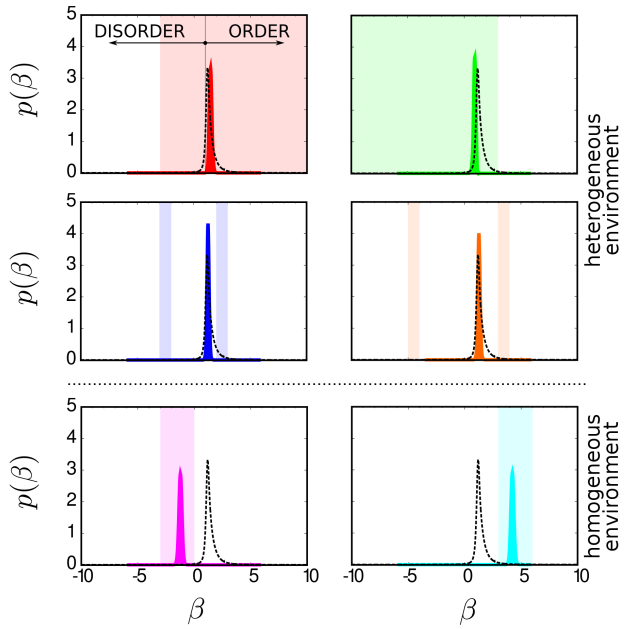


Figure 28: Evolutionary model leading to near to criticality in complex environments. A community of M agents undergoes a genetic algorithm dynamics [120]. Each agent is simultaneously exposed to diverse stimuli $\mathbf{s}$ provided by S different sources, each one characterized by a probability $P_{src}(\mathbf{s}|\alpha^u)$ with $u = 1, \ldots, S$, fully specified by parameters α^u. At each time step, S sources are randomly drawn with probability $\rho_{src}(\alpha^u)$ (in this case a uniform distribution with support in the colored region). Each agent i ($i = 1, ..., M$) has an internal state $P_{int}(\mathbf{s}|\beta^i)$ aimed at representing –or coping with– the environment. Agents' fitness increases as the mean Kullback-Leibler divergence from the set of sources to which they are exposed decreases. An agent that is killed is replaced by a new individual with a parameter β inherited from one of the other agents (and, with some probability, a small variation/mutation). The community dynamically evolves and eventually reaches a steady state distribution of parameters, $p(\beta)$. The six panels in the figure correspond to different supports (colored regions) for uniform source distributions, $\rho_{src}(\alpha^u)$. The dashed line is the generalized susceptibility (Fisher information) of the internal probability distribution, which exhibits a peak at the critical point separating an ordered from a disordered phase. Heterogeneous source pools (upper and middle panels) lead to distributions peaked at criticality, whereas for homogeneous sources (lower panels), the communities are not critical but specialized. Stimuli distributions are parametrized in a rather simplistic way as $P_{src}(\mathbf{s}|\alpha^u) \propto \exp\{\alpha^u \frac{N}{2}(\sum_{k=1}^{N} \frac{s_k}{N})^2\}$, while internal states are identical but replacing α^u by β^i. In the guiding example of genetic regulatory networks, this example corresponds to an extremely simple fully-connected network in which the state of each gene is equally determined by all the other genes and, hence, the probability of a given state depends only on the total number of on/off genes, controlled by a single parameter.

For sufficiently large values of N, the sum in eq. 124 cannot be explicitly computed because of the diverging large number of states, 2^N. However, since equations 115 and 114 depend on $\mathbf{s}$ only through the magnetization $m = \sum_i s_i/N$,

$$H_{src}(\mathbf{s}|\boldsymbol{\alpha}) = -\frac{N}{2}\alpha_1 m^2 \tag{126}$$

and

$$H_{int}(\mathbf{s}|\boldsymbol{\beta}) = -\frac{N}{2}\beta_1 m^2, \tag{127}$$

we can compute the sum as follows. Defining $\Gamma(m)$ as the number of states with $\sum_i s_i/N = m$, and using the Stirling approximation, one readily obtains:

$$\Gamma(m) = \binom{N}{\frac{N(1+m)}{2}} \underset{N\gg 1}{=} \exp\left\{-N\left(\frac{1+m}{2}\log\frac{1+m}{2} + \frac{1-m}{2}\log\frac{1-m}{2}\right)\right\}. \tag{128}$$

Then, the KL divergence can be computed as

$$D(\boldsymbol{\alpha}|\boldsymbol{\beta}) \underset{N\gg 1}{=} \int_{-1}^{1} \Gamma(m)\hat{P}_{src}(m|\boldsymbol{\alpha}) \log\frac{\hat{P}_{src}(m|\boldsymbol{\alpha})}{\hat{P}_{int}(m|\boldsymbol{\beta})}\,dm. \tag{129}$$

The same type of approximation can be used to calculate the normalization of P_{src} and P_{int}.

6.5.0.1 *Results*

Having computed the KL as a function of parameter values we can iterate the evolutionary dynamics, as described before. Figure 29 illustrates that starting from different initial conditions $p(\boldsymbol{\beta}, 0)$ after some (sufficiently long) times the ensemble of individuals converges to a unique steady state $p(\boldsymbol{\beta}, t \to \infty)$. The resulting distribution is sharply peaked very near the critical point, at the very same location at which the Fisher information or generalized susceptibility peaks. This peak approaches the critical point $\beta = \beta_c$ in the limit $N \to \infty$.

First, we proceed to analyze the relaxation of the initial distribution of parameters $p(\boldsymbol{\beta}, t = 0)$ at the stationary one $p(\boldsymbol{\beta})$, and its dependence on the initial condition. These results are plotted in Fig. 29. Fig. 30 illustrates the dependence of the results on parameters.

The main conclusions are:

- The system becomes closer and closer to the critical point as the system-size N is enlarged.
- The distribution reaches an asymptotic shape as the ensemble size grows.
- The distribution becomes sharper for smaller mutation rates.

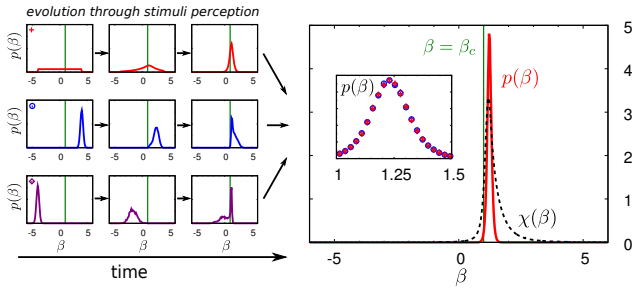


Figure 29: *Time evolution in the Evolutionary Model* (with parametrization of eq. 114 with $\alpha_2 = 0$ and eq. 115): Panels on the left represent the evolution of three different initial distributions of the parameter β in the community. In all cases the environment is described by the uniform distribution of parameters $\rho_{src}(\alpha) = U([-10, 10])$. The distributions converge to the same stationary state –points and red line on the right panel–, which is peaked at the maximum of the generalized susceptibility (dashed line curve). The red line corresponds to an initial uniform distribution $U([-4, 4])$, and blue and purple lines to Gaussian distributions $N(4, 0.25)$ and $N(-4, 0.25)$, respectively. Parameters are the same as in Table 2, and $R = 10^4$ independent realizations.

- The distribution becomes much sharper for small mutation variances.
- The distribution reaches an asymptotic shape as the number of external sources is increased.

6.6 EFFECTIVE CRITICALITY OF THE ENVIRONMENT

As we have shown in section 6.5, minimizing the KL divergence to the "averaged environment" –i.e. the distribution of sources resulting by averaging over different environmental parameters– leads to the same result as minimizing the mean KL divergence to the sources, which is what we implement in the simulations. Therefore, agents seeing a complex "averaged environment" tend to become critical.

It has been recently suggested that marginalizing a distribution over some parameters can generally lead to effective critical distributions [97]. Thus it may not be surprising that individuals tune their parameters near the criticality to minimize the KL divergence with respect to such a critical environment.

Here we show that our results cannot be generically explained in terms of this phenomenon. For the cases in which the individuals end up near criticality, the averaged environment is not necessarily critical. To illustrate this, we consider each of the pools of sources used in Fig. 3 of the main text to illustrate the evolution of agents in the presence of complex environments. As it is not possible to identify the criticality of the "average environment" by looking at the peak of

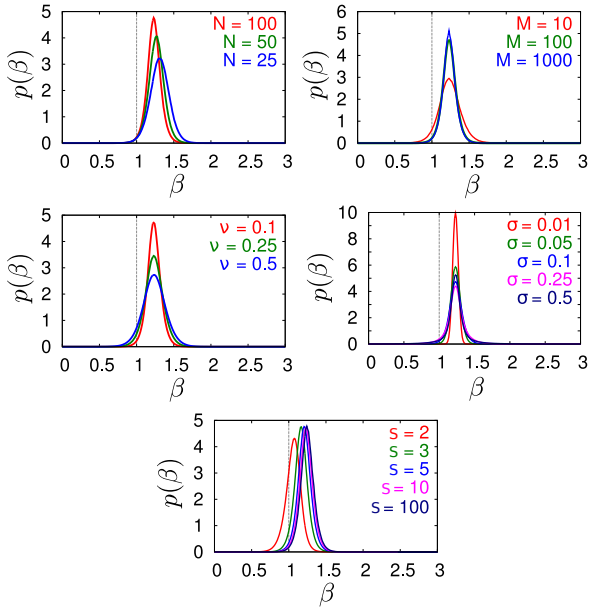


Figure 30: *Dependence on parameters in the Evolutionary Model:* Stationary distribution $p(\beta)$ as a function of diverse parameters; different colors in each plot stand for different values of (from top to bottom and from left to right): N, community size M, mutation probability ν, mutation deviation σ, and number of external sources S. Unless otherwise stated, other parameters take the same values as in Table 2. The dashed lines indicate the critical point location (in the limit $N \to \infty$)

a susceptibility, following the seminal paper by Mora and Bialek[76] we say that a particular distribution is critical if it obeys Zipf's law. To check this, for each particular environment, we plot the probability of states ordered by their rank and the energy as a function of the entropy (see [76]). At the critical point, the energy should be a linear function of the entropy, and the rank ordering should obey the Zipf's law [76, 97].

The results are shown in Fig. 31, where we have kept the same relative position and color code as in the original plot (Fig.3 of the main plot). Only two of the six averaged environments presented in Fig. 3 of the main text turn out to be critical in the sense of Zipf's law (upper panels in Fig. 31). Furthermore, in the two central cases, the averaged environment is not critical, but the optimal internal distribution peaks around criticality (right panels in Fig. 31). This case corresponds to an environment composed, essentially, of two very different type of sources, and individuals have to accommodate to the critical point to respond to both of them efficiently. This demonstrates that our approach works in a general scenario of heterogeneous environmental sources, without requiring the environment to be itself critical.

6.7 DISCUSSION AND CONCLUSIONS

Under the mild assumption that living systems need to construct good though approximate internal representations of the outer complex world and that such representations are encoded in terms of probability distributions, we have shown –by employing concepts from statistical mechanics and information theory– that the encoding probability distributions do necessarily lie where the *generalized susceptibility* or *Fisher information* exhibits a peak [113], i.e. in the vicinity of a critical point, providing the best possible compromise to accommodate both regular and noisy signals.

In the presence of broadly different ever-changing heterogeneous environments, computational evolutionary and adaptive models vividly illustrate how a collection of living systems eventually cluster near the critical state. A more accurate convergence to criticality is found in a co-evolutionary/co-adaptive set-up in which individuals evolve/adapt to represent with fidelity other agents in the community, thereby creating a collective "language", which turns out to be critical.

These ideas apply straightforwardly to genetic and neural networks –where they could contribute to a better understanding of why neural activity seems to be tuned to criticality– but have a broader range of implications for general complex adaptive systems [107]. For example, our framework could be applicable to some bacterial communities for which a huge phenotypic (internal state) variability has been empirically observed [121]. Such a large phenotypic diversification

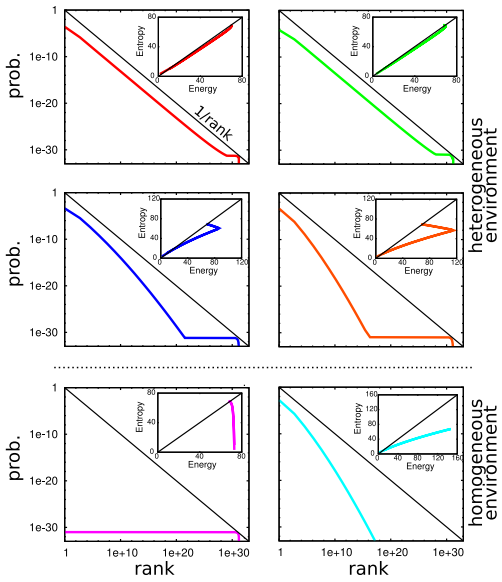


Figure 31: *Heterogeneity and criticality of the averaged environment.* Each panel of the figure refers to the corresponding panel, with the same position and color code, in Fig. 3 of the main text. In each panel, the main plot shows the probability of states in the "averaged environment" $P_{env}(\mathbf{s}|\rho_{src})$ with the states $\mathbf{s}$ ranked in order of decreasing probability. The inset shows the energy associated to $\bar{P}_{env}$ as a function of the entropy. The black lines define the expected linear behavior in the critical case [76]. Only the red and green settings correspond to a critical distribution obeying also Zipf's law ($1/x$). The more interesting cases are the blue and the orange ones: for these, the internal distribution is critical even though the average environment is not.

can be seen as a form of "bet hedging", an adaptive survival strategy analogous to stock-market portfolio management [122], which turns out to be a straightforward consequence of individuals in the community being critical. Usually, from this point of view, generic networks diversify their "assets" among multiple phenotypes to minimize the long-term risk of extinction and maximize the long-term expected growth rate in the presence of environmental uncertainty [122]. Similar bet-hedging strategies have been detected in viral populations and could be explained as a consequence of their respective communities having converged to a critical state, maximizing the hedging effect. Similarly, criticality has been recently shown to emerge through adaptive information processing in machine learning, where networks are trained to produce a desired output from a given input in a noisy environment; when tasks of very different complexity need to be simultaneously learned, networks adapt to a critical state to enhance their performance [123]. In summary, criticality in some living systems could result from the interplay between their need for producing accurate representations of the world, their need to cope with many widely diverse environmental conditions, and their well-honed ability to react to external changes in an efficient way. Evolution and adaptation might drive living systems to criticality in response to this smart cartography.

7 CONCLUSION AND PERSPECTIVES

Si finis bonus est, totum bonum erit.

— Gestæ Romanorum, Tale LXVII

In these thesis we have discussed two main problems, both related with biological interactions.

In the first part we have shown how random matrix theory can be extremely useful to model biological interactions. While most of ecological models are focused on the details and consider few species at a time, this method start from an equivalent of the thermodynamic limit and identify the effective order parameters determining the stability of ecosystems. The main result is the quantification of the role of network structure and interaction strengths in determining stability or the persistence in a randomly fragmented landscapes.

The second part was dedicated to the fact that many phenomena observed close to a critical point are seen in living systems. We have introduced a model based on the needs of individuals of building a reliable representation of the external world, that explains how critical phenomena can emerge without fine-tuning any parameter.

What to do next? There are several theoretical problems that stem from both the two parts, but probably the most important perspectives are the experimental ones. In both the cases, natural systems are too complicated and too uncontrollable to be a good test for the model and the methods explained in this thesis. For instance it is impossible to control the perturbations in a natural food-web in order to study the effect of the structure or the path leading to instability. In a very similar way, one cannot have a control on the birds of a flock and tune their interactions and see the effect on the system. Having controllable systems in necessarily in both the cases.

Part III

APPENDIX

A

FROM STOCHASTIC EQUATION TO ASYMPTOTIC LINEAR STABILITY

A.1 DETERMINISTIC EQUATION

We know (almost) everything about the deterministic generalized Lotka-Volterra equation. If we have a community of S species, the evolution of their population is defined by

$$\partial_t n_i = n_i \left(r_i + \sum_j I_{ij}(\mathbf{n}) n_j \right) , \tag{130}$$

where $I(\mathbf{n})$ is the interaction matrix which may also depend on the populations $\mathbf{n}$ to take into account for non-linear functional responses (e.g. Holling type II). Very generally we can disentangle three terms contributing to $I(\mathbf{n})$: a non-negative matrix $M(\mathbf{n})$ (mutualistic interactions), a non-positive matrix $C(\mathbf{n})$ (competitive interactions) and a (sign)-skew-symmetric matrix $P(\mathbf{n})$ (predator-prey interactions). We also impose the ecological (obvious?) condition that given a pair of species i and j at most one of the three elements M_{ij}, C_{ij} and P_{ij} is non zero (i.e. the species i cannot interact in two different ways with the species j). The self-interaction is a competition interaction, if one does not consider any exotic self-interaction (e.g. cannibalism), therefore $C_{ii} = 1/K_i$ and $M_{ii} = P_{ii} = 0$.

A.2 STOCHASTIC EQUATION

Let us consider a very general multi-species birth-death process

$$\frac{\partial P(\mathbf{n}, t)}{\partial t} = \sum_i \left[(E_i^- - 1) D_i(\mathbf{n}) P(\mathbf{n}, t) + (E_i^+ - 1) B_i(\mathbf{n}) P(\mathbf{n}, t) \right] , \tag{131}$$

where the operators $E_i^{\pm}$ act as explained in [124]. There are several ways to define the rates $D_i(\mathbf{n})$ and $B_i(\mathbf{n})$ and obtain equation 138 as deterministic limit. One way is to define some "chemical reactions" [124]. We consider a simpler way. The deterministic equation corresponding to eq. 131 is

$$\partial_t \langle n \rangle_i = B_i(\langle \mathbf{n} \rangle) - D_i(\langle \mathbf{n} \rangle) := R_i(\langle \mathbf{n} \rangle) . \tag{132}$$

If one consider that $\mathbf{n}$ can take values in $\mathbf{Z}^n$ (this one is the real difference with the "chemical reaction" approach), there is a very natural way to define $B_i(\mathbf{n})$ and $D_i(\mathbf{n})$

$$\begin{aligned} B_i(\mathbf{n}) &= n_i\left(b_i + \sum_j M_{ij}n_j + \sum_j \Theta(P_{ij})P_{ij}n_j\right) \\ D_i(\mathbf{n}) &= n_i\left(d_i - \textstyle\sum_j C_{ij}n_j - \sum_j \Theta(-P_{ij})P_{ij}n_j\right) . \end{aligned} \tag{133}$$

The only arbitrary things that cannot be measured by a deterministic analysis are the rates b_i and d_i, given the fact that in the deterministic equation it appears only $r_i = b_i - d_i$. Now we have all the ingredients to play with.

A.3 KRAMERS-MOYAL EXPANSION

Master equations are beautiful but analytically intractable. To get something tractable we have to make some approximations. The simplest things to do is to approximate equation 131 with the Kramers-Moyal expansion, promoting $\mathbf{n}$ to be a continous variable $\mathbf{y}$, obtaining the following Fokker-Plank equation

$$\frac{\partial P(\mathbf{y},t)}{\partial t} = \sum_i \left\{ \frac{\partial}{\partial y_i}\left[\Big(R_i(\mathbf{y})\Big)P(\mathbf{y},t)\right] + \frac{1}{2}\frac{\partial^2}{\partial y_i^2}\left[\Big(F_i(\mathbf{y})\Big)P(\mathbf{y},t)\right]\right\} , \tag{134}$$

where $R_i(\mathbf{y}) = B_i(\mathbf{y}) - D_i(\mathbf{y})$ and $F_i(\mathbf{y}) = B_i(\mathbf{y}) + D_i(\mathbf{y})$.

A.4 VAN KAMPEN'S ANSATZ

We will consider the system size expansion [125, 126] (SSE) which is typically very useful because it allows analytical considerations. What it is crucial in the SSE is the Van Kampen's ansatz

$$y_i = y_i^* + x_i , \tag{135}$$

where y_i^* is the average population (i.e. the solution of the deterministic equations $B_i(\mathbf{y}^*) = D_i(\mathbf{y}^*)$), while x_i are fluctuations around the average (of order $\sqrt{y_i^*}$). The strong assumption is that the fluctuations are small around the average. The main drawback of this approach is therefore that one cannot study anything which is not fluctuation around the average (e.g. one does not see the trajectories to the absorbing points). In any case SSE returns much more informations than the ones obtained with the deterministic analysis.

The SSE, once applied to eq.131, returns two equations, one defines the evolution of y_i^* (i.e. is the deterministic equation, in our case eq. 138), while the other one is a Fokker-Plank equations for

the fluctuations x_i, which correspond to the following Langevin (Ito prescription)

$$\partial_t x_i = \sum_j A_{ij} x_j + \sqrt{Q_i}\xi_i \,, \tag{136}$$

where ξ_i are white uncorrelated noises (i.e. $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t)\xi_j(t') \rangle = \delta_{ij}\delta(t-t')$), while

$$\begin{aligned} Q_i &:= F_i(\mathbf{y}^*) = 2D_i(\mathbf{y}^*) \\ A_{ij} &:= \left.\frac{\partial R_i(\mathbf{n})}{\partial n_j}\right|_{\mathbf{y}^*} \,, \end{aligned} \tag{137}$$

where $\mathbf{y}^*$ is the solution of the stationary deterministic equation, i.e.

$$B_i(\mathbf{y}^*) = D_i(\mathbf{y}^*) \,. \tag{138}$$

Note also that A_{ij} is the Jacobian evaluated at the fixed point, i.e. it is exactly the community matrix studied by May.

A.5 VAN KAMPEN FEAT. MAY

As we are interested in the fluctuations around the the fixed point, we can study eq 136. By applying the Fourier transform, we obtain

$$i\omega \widehat{x}_i = \sum_j A_{ij}\widehat{x}_j + \sqrt{Q_i}\widehat{\xi}_i \,, \tag{139}$$

obtaining

$$\widehat{x}_i(\omega) = \sum_j (i\omega - A)^{-1}_{ij} \sqrt{Q_j}\widehat{\xi}_j(\omega) \,. \tag{140}$$

One can write the solution of equation 140 as

$$\widehat{x}_i(\omega) = \sum_{jk} \frac{U_{ik}U^{-1}_{kj}}{i\omega - \lambda_k} \sqrt{Q_j}\widehat{\xi}_j(\omega) \,, \tag{141}$$

where $U^{-1}_{ik}A_{ks}U_{sj} = \lambda_i\delta_{ij}$. Note that we can aplly the SSE only if the fixed point (the deterministic solution) is stable, otherwise the fluctuation would not be small. Therefore all the eigenvalues λ_k have a negative real part.

We can immediately obtain the correlation matrix

$$\begin{aligned} \widehat{C}_{ij}(\omega) := \langle \widehat{x}_i(\omega)\widehat{x}_j(-\omega) \rangle = \sum_{skl} \frac{U_{ik}U^{-1}_{ks}U_{jl}U^{-1}_{ls}}{(\lambda_k - i\omega)(\lambda_l + i\omega)} Q_s = \\ \sum_{kl} \frac{U_{ik}U_{jl}}{(\lambda_k - i\omega)(\lambda_l + i\omega)} F_{kl} \,, \end{aligned} \tag{142}$$

where $F_{kl} = \sum_s \left[U^{-1}_{ks} U^{-1}_{ls} Q_s \right]$. The power-spectrum is simply

$$S_j(\omega) = \widehat{C}_{jj}(\omega) = \sum_{kl} \frac{U_{jk} U_{jl}}{(\lambda_k - i\omega)(\lambda_l + i\omega)} F_{kl} \, . \tag{143}$$

From eq. 142 we can obtain the temporal correlation by calculating

$$C_{ij}(t) := \left\langle x_i(0) x_j(t) \right\rangle = \int d\omega \, \frac{e^{-i\omega t}}{2\pi} \left\langle \widehat{x}_i(\omega) \widehat{x}_j(-\omega) \right\rangle \, , \tag{144}$$

and therefore

$$\begin{aligned} C_{ij}(t) = \sum_{kl} U_{ik} U_{jl} F_{kl} \int d\omega \, \frac{e^{-i\omega t}}{2\pi(\lambda_k - i\omega)(\lambda_l + i\omega)} = \\ \sum_{kl} U_{ik} U_{jl} F_{kl} \left[\frac{e^{\lambda_k t} + e^{\lambda_l t}}{\lambda_k + \lambda_l} \right] . \end{aligned} \tag{145}$$

B

DEMONSTRATION OF WIGNER SEMI-CIRCULAR LAW

The matrix ensemble distribution is defined over random matrices with the following density

$$P(M) = \frac{1}{Z} \exp(-\frac{S}{4} \operatorname{tr}(M^2)) \,, \tag{146}$$

The distribution of eigenvalues is defined as

$$\rho(z) := \langle \sum_i \delta(z - \lambda_i) \rangle \,, \tag{147}$$

where the average is calculated over the distribution of matrices of equation 146.

Introducing $\delta(x) = \lim_{\epsilon \to 0^+} 1/(x + i\epsilon)$, one has

$$\rho(z) = \lim_{\epsilon \to 0^+} \Im \langle G_\epsilon(z) \rangle \,, \tag{148}$$

where the resolvent is defined as

$$G_\epsilon(z) := \frac{1}{N} \operatorname{tr} \frac{1}{z - M + i\epsilon} \,. \tag{149}$$

We introduce the partition function

$$Z(J) := \langle \frac{\det(z - M + J)}{\det(z - M)} \rangle \,, \tag{150}$$

from which it is possible to obtain the resolvent by deriving over J and dividing by N.

We can write both the determinants that appear in eq. 150 as two integrals over anti-commuting (Grassman) variables and ordinary commuting variables

$$\det(z - M + J) := \int \prod_{i=1}^{N} d\chi_k d\chi_k^* \exp\left(-i \sum_{kl} \chi_k^* (z\delta_{kl} - J\delta_{kl} + M_{kl}) \chi_l \right) \,, \tag{151}$$

where χ are Grassman variables. The other term will be

$$\frac{1}{\det(z - M)} := \int \prod_{i=1}^{N} \frac{i}{2\pi} d\phi_k d\phi_k^* \exp\left(-i \sum_{kl} \phi_k^* (z\delta_{kl} + M_{kl}) \phi_l \right) \,. \tag{152}$$

Combining the two terms, we obtain

$$Z(J) = \int [d\Psi] \exp\Big(-i \sum_k (z\phi_k^* \phi_k + (z+J)\chi_k^* \chi_k)\Big) \int [dM] \exp\Big(-\frac{N}{2} \sum_{jk} M_{jk}^2\Big) \exp\Big(i \sum_{jk} M_{jk} (\phi_j^* \phi_k + \chi_j^* \chi_l)\Big) \,, \tag{153}$$

where $[d\Psi] = \prod_{j=1}^{N} \frac{i}{2\pi} d\phi_k^* d\phi_k d\chi_k^* d\chi_k$ and $[dM] = \prod_{jk} M_{jk} \delta(M_{jk} - M_{kj})$. By integrating over M we obtain

$$Z(J) = \int [d\Psi] \exp\Big(-i \sum_k (z\phi_k^* \phi_k + (z+J)\chi_k^* \chi_k)\Big) \exp\Big(-\frac{1}{2N} \sum_{jk} (\phi_j^* \phi_k + \chi_j^* \chi_k)^2\Big) \,. \tag{154}$$

The last exponential is quartic in the fields. One can rewrite it in the following way

$$\sum_{jk} (\phi_j^* \phi_k + \chi_j^* \chi_k)^2 = \operatorname{trg} \begin{pmatrix} \sum_j \phi_j^* \phi_j & \sum_j \chi_j^* \phi_j \\ \sum_j \chi_j \phi_j^* & \sum_j \chi_j^* \chi_j \end{pmatrix}^2 =: \operatorname{trg} \Gamma^2 \,, \tag{155}$$

where trg is the supersymmetric trace [127]. Note that Γ is a supermatrix, i.e. its diagonal elements are commuting variables, while off-diagonal elements are anticommuting. The partition function is therefore

$$Z(J) = \int [d\Psi] \exp\Big(-\frac{1}{2N} \operatorname{trg} \Gamma^2 - i \sum_{jk} \Psi_j^* \zeta \Psi_j\Big) \,, \tag{156}$$

where we have introduced the superfield $\Phi_j = (\phi_j, \chi_j)^t$ and the 2×2 matrix $\zeta_{ij} = (z + J\delta_{i2})\delta_{ij}$.

The argument in the exponential of equation 156 is quartic in the fields. In order to make it quadratic we can introduce a Hubbard-Stratonovich transformation

$$\exp\Big(-\frac{1}{2N} \operatorname{trg} \Gamma^2\Big) = \int [d\sigma] \exp\Big(-\frac{N}{2} \operatorname{trg} \sigma^2 - i \sum_{jk} \Psi_j^* \sigma \Psi_k\Big) \,, \tag{157}$$

where σ in a supermatrix such that

$$\sum_{jk} (\phi_j^* \phi_k + \chi_j^* \chi_k)^2 = \operatorname{trg} \begin{pmatrix} a & \alpha \\ \beta & ib \end{pmatrix} \,, \tag{158}$$

where a and b are commuting variables, while α and β are Grassman variables. We obtain

$$Z(J) = \int [d\Psi][d\sigma] \exp\Big(-\frac{N}{2} \operatorname{trg} \sigma^2 - i \sum_{jk} \Psi_j^* (\sigma + \zeta) \Psi_k\Big) \,. \tag{159}$$

We can now integrate over the fields Ψ, since it is a Gaussian integral. We obtain

$$Z(J) = \int [d\sigma] \exp\left(-\frac{N}{2} \operatorname{trg} \sigma^2 - N \operatorname{trg} \log(\sigma + \zeta)\right) . \tag{160}$$

By introducing the parametrization of σ shown in equation 158 we obtain

$$Z(J) = \frac{1}{2\pi} \int da db d\alpha d\beta \, \frac{(ib + J + \zeta - \frac{\alpha\beta}{a+z})^N}{(a+z)^N} \exp\left(-\frac{N}{2}(a^2 + b^2 + 2\alpha\beta)\right) . \tag{161}$$

We can make two change of variables $x = a + z$ and $y = b - iJ - iz$, obtaing

$$Z(J) = \frac{1}{2\pi} \int dx dy d\alpha d\beta \, (\frac{iy}{x} - \frac{\alpha\beta}{x^2})^N \exp\left(-\frac{N}{2}((x-z)^2 + (y + iz + iJ)^2 + 2\alpha\beta)\right) . \tag{162}$$

We can expand the Grassmann variables, obtaining

$$Z(J) = \frac{1}{2\pi} \int dx dy d\alpha d\beta \, (\frac{iy}{x})^N (1 - N\frac{\alpha\beta}{x})(1 - N\alpha\beta) \exp\left(-\frac{N}{2}((x-z)^2 + (y + iz + iJ)^2)\right) , \tag{163}$$

and we can finally integrate over α and β, obtaining

$$Z(J) = \frac{1}{2\pi} \int dx dy \, \frac{iy}{x})^N (1 + \frac{1}{ixy}) \exp\left(-\frac{N}{2}((x-z)^2 + (y + iz + iJ)^2)\right) . \tag{164}$$

At this point it is possible to perform the integration over x and y and obtain an expression for eigenvalues density at any N [127]. Since we are interested at the large N behavior, we use the saddle point method. We can write the resolvent as

$$\langle G_\epsilon(z) \rangle := \frac{dZ(j)}{dj}\Big|_{J=0} = \frac{1}{2\pi} \int dx dy \, (z - iy) \exp(-NS(x, y, z, J)) , \tag{165}$$

where the leading term of S is

$$S(x, y, z) = -\log(y/x) + \frac{1}{2}((x-z)^2 + (y + iz)^2)) . \tag{166}$$

We can then minimize S with respect to y and x, obtaining that the minimum is given by the solution of $1/y_m = y_m + iz$ and $1/x_m + x_m - z = 0$. Introducing this in equation 165, we obtain

$$\langle G_\epsilon(z) \rangle = z - iy_m = \frac{z}{2} - \frac{1}{2}\sqrt{4 - z^2} , \tag{167}$$

and using the definition of equation 148 we finally get

$$\rho(z) = \frac{1}{2\pi}\sqrt{4 - z^2} , \tag{168}$$

which coincides with Wigner's solution.

BIBLIOGRAPHY

[1] Galileo Galilei. *Il Saggiatore (The Assayer)*. Doubleday, New York, 1957.

[2] E Schrodinger. *What is Life?: The Physical Aspects of Living Cell with Mind and Matter and Autobiographical Sketches*. Cambridge University Press, 1967.

[3] William Bialek. *Biophysics: searching for principles*. Princeton University Press, 2012.

[4] Stephen Jay Gould. *{H}en's {T}eeth and {H}orse's {T}oes*. WW Norton \& Company, 1983.

[5] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[6] M E J Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23):8577–82, June 2006.

[7] M. Newman. Assortative Mixing in Networks. *Physical Review Letters*, 89(20):208701, October 2002.

[8] Jordi Bascompte, Pedro Jordano, Carlos J Melián, and Jens M Olesen. The nested assembly of plant-animal mutualistic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 100(16):9383–7, August 2003.

[9] Jens M Olesen, Jordi Bascompte, Yoko L Dupont, and Pedro Jordano. The modularity of pollination networks. *Proceedings of the National Academy of Sciences of the United States of America*, 104(50):19891–6, December 2007.

[10] Mário Almeida-Neto, Paulo Guimarães, Paulo R. Guimarães, Rafael D. Loyola, and Werner Ulrich. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, 117(8):1227–1239, August 2008.

[11] Phillip P A Staniczenko, Jason C Kopp, and Stefano Allesina. The ghost of nestedness in ecological networks. *Nature communications*, 4:1391, January 2013.

[12] Samuel Jonhson, Virginia Domínguez-García, and Miguel A Muñoz. Factors determining nestedness in complex networks. *PloS one*, 8(9):e74025, January 2013.

[13] Samuel Johnson, Joaquín J. Torres, J. Marro, and Miguel A. Muñoz. Entropic Origin of Disassortativity in Complex Networks. *Physical Review Letters*, 104(10):108702, March 2010.

[14] Miguel A Fortuna and Jordi Bascompte. Habitat loss and the structure of plant-animal mutualistic networks. *Ecology letters*, 9(3):281–6, March 2006.

[15] Abhay Krishna, Paulo R. Guimarães Jr, Pedro Jordano, and Jordi Bascompte. A neutral-niche theory of nestedness in mutualistic networks. *Oikos*, 117(11):1609–1618, October 2008.

[16] Ugo Bastolla, Miguel A Fortuna, Alberto Pascual-García, Antonio Ferrera, Bartolo Luque, and Jordi Bascompte. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, 458(7241):1018–20, April 2009.

[17] Elisa Thébault and Colin Fontaine. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science (New York, N.Y.)*, 329(5993):853–6, August 2010.

[18] Samir Suweis, Filippo Simini, Jayanth R Banavar, and Amos Maritan. Emergence of structural and dynamical properties of ecological mutualistic networks. *Nature*, 500(7463):449–52, August 2013.

[19] Jennifer A Dunne, Richard J Williams, and Neo D Martinez. Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12917–22, October 2002.

[20] R Milo, S Shen-Orr, S Itzkovitz, N Kashtan, D Chklovskii, and U Alon. Network motifs: simple building blocks of complex networks. *Science (New York, N.Y.)*, 298(5594):824–7, October 2002.

[21] Diego Garlaschelli, Guido Caldarelli, and Luciano Pietronero. Universal scaling relations in food webs. *Nature*, 423(6936):165–8, May 2003.

[22] J E Cohen, F Briand, and C M Newman. *Community food webs: Data and theory*. Springer-Verlag, Berlin, 1990.

[23] R J Williams and N D Martinez. Simple rules yield complex food webs. *Nature*, 404(6774):180–3, March 2000.

[24] Stefano Allesina, David Alonso, and Mercedes Pascual. A general model for food web structure. *science*, 320(5876):658–661, May 2008.

[25] Robert M May. Will a Large Complex System be Stable? *Nature*, 238(5364):413–414, August 1972.

[26] Terence Tao and Van Vu. Random Matrices: Universality of Local Eigenvalue Statistics up to the Edge. *Communications in Mathematical Physics*, 298(2):549–572, April 2010.

[27] Terence Tao and Van Vu. Random matrices: Universality of local eigenvalue statistics. *Acta Mathematica*, 206(1):127–204, March 2011.

[28] H. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein. Spectrum of Large Random Asymmetric Matrices. *Physical Review Letters*, 60(19):1895–1898, May 1988.

[29] J. J. M. Verbaarschot. The Supersymmetric Method in Random Matrix Theory and Applications to QCD. October 2004.

[30] Alexandru Nica and Roland Speicher. *Lectures on the Combinatorics of Free Probability, Volume 13*. Cambridge University Press, 2006.

[31] Vyacheslav L Girko. Circular law. *Teoriya Veroyatnostei I Ee Primeneniya*, 29(4):669–679, 1984.

[32] Stefano Allesina and Si Tang. Stability criteria for complex ecosystems. *Nature*, February 2012.

[33] Samir Suweis, Jacopo Grilli, and Amos Maritan. Disentangling the effect of hybrid interactions and of the constant effort hypothesis on ecological community stability. *Oikos*, 123(5):525–532, May 2014.

[34] Si Tang, Samraat Pawar, and Stefano Allesina. Correlation between interaction strengths drives stability in large ecological networks. *Ecology letters*, 17(9):1094–1100, 2014.

[35] A Mougi and M Kondoh. Diversity of interaction types and ecological community stability. *Science (New York, N.Y.)*, 337(6092):349–51, July 2012.

[36] Aldous Huxley. *Brave New World*. Harper Perennial Modern Classics, 2006.

[37] Serguei Saavedra, Rudolf P Rohr, Vasilis Dakos, and Jordi Bascompte. Estimating the tolerance of species to the effects of global environmental change. *Nature communications*, 4:2350, January 2013.

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

[39] Leslie Hogben. *Handbook of Linear Algebra*, volume 2. CRC Press, 2006.

[40] D Hershkowitz. Positivity of principal minors, sign symmetry and stability. *Linear Algebra and its Applications*, 364:105–124, May 2003.

[41] Si Tang and Stefano Allesina. Reactivity and stability of large ecosystems. *Frontiers in Ecology and Evolution*, 2, June 2014.

[42] Giovanni Strona, Domenico Nappo, Francesco Boccacci, Simone Fattorini, and Jesus San-Miguel-Ayanz. A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nature communications*, 5:4114, January 2014.

[43] Eric Hoffer. Absolute Faith Corrupts Absolutely. *The New York Times Magazine*, April 25:25, 1971.

[44] Ilkka Hanski. *Metapopulation ecology*. Oxford University Press, Oxford New York, 1999.

[45] I Hanski and O Ovaskainen. The metapopulation capacity of a fragmented landscape. *Nature*, 404:755–758, 2000.

[46] O Ovaskainen and I Hanski. Spatially Structured Metapopulation Models: Global and Local Assessment of Metapopulation Capacity. *Theoretical Population Biology*, 60:281–302, 2001.

[47] I Hanski and O Ovaskainen. Metapopulation theory for fragmented landscapes. *Theoretical Population Biology*, 64:119–127, 2003.

[48] R Levins. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, 15:237–240, 1969.

[49] Iwan Jensen. Critical behavior of the pair contact process. *Physical Review Letters*, 70(10):1465–1468, March 1993.

[50] M Mézard, G Parisi, and A Zee. Spectra of Euclidean random matrices. *Nuclear Physics B*, 559(3):689–701, 1999.

[51] Marc Mézard and Giorgio Parisi. Thermodynamics of Glasses: A First Principles Computation. *Phys. Rev. Lett.*, 82(4):747–750, 1999.

[52] Ten Ming Wu and Roger F Loring. No Title.

[53] M Penrose. *Random Geometric Graphs*. Oxford University Press, New York, 2003.

[54] H Caswell. *Matrix population models: {C}onstruction, analysis and interpretation. 2nd edition*. Sinauer Associates, 2001.

[55] S P Otto and T Day. *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*. Princeton University Press, Princeton, USA, 2007.

[56] S G Walker and P Van Mieghem. On lower bounds for the largest eigenvalue of a symmetric matrix. *Linear Algebra and its Applications*, 429:519–526, 2008.

[57] Luis J Gilarranz and Jordi Bascompte. Spatial network structure and metapopulation persistence. *Journal of theoretical biology*, 297:11–16, 2012.

[58] Dominique Gravel, Timothée Poisot, and Philippe Desjardins-Proulx. Using neutral theory to reveal the contribution of meta-community processes to assembly in complex landscapes. *Journal of Limnology*, 73(s1), 2014.

[59] Jesper Dall and Michael Christensen. Random geometric graphs. *Physical Review E*, 66(1):16121, 2002.

[60] Louis W Botsford, Alan Hastings, and Steven D Gaines. Dependence of sustainability on the configuration of marine reserves and larval dispersal distance. *Ecology Letters*, 4(2):144–150, 2001.

[61] Steven D Gaines, Brian Gaylord, and John L Largier. Avoiding current oversights in marine reserve design. *Ecological Applications*, 13(sp1):32–46, 2003.

[62] David M Kaplan and Louis W Botsford. Effects of variability in spacing of coastal marine reserves on fisheries yield and sustainability. *Canadian Journal of Fisheries and Aquatic Sciences*, 62(4):905–912, 2005.

[63] Otso Ovaskainen. Habitat destruction, habitat restoration and eigenvector–eigenvalue relations. *Mathematical biosciences*, 181(2):165–176, 2003.

[64] Thomas Guhr, Axel Müller-Groeling, and Hans A Weidenmüller. Random-matrix theories in quantum physics: common concepts. *Physics Reports*, 299(4):189–425, 1998.

[65] Patrick A P Moran. Notes on continuous stochastic phenomena. *Biometrika*, 37(1-2):17–23, 1950.

[66] A G Bunn, D L Urban, and T H Keitt. Landscape connectivity: a conservation application of graph theory. *Journal of Environmental Management*, 59(4):265–278, 2000.

[67] Dean L Urban, Emily S Minor, Eric A Treml, and Robert S Schick. Graph models of habitat mosaics. *Ecology letters*, 12(3):260–273, 2009.

[68] Stephen J Cornell and Otso Ovaskainen. Exact asymptotic analysis for metapopulation dynamics on correlated dynamic landscapes. *Theoretical population biology*, 74(3):209–225, 2008.

[69] Ace North and Otso Ovaskainen. Interactions between dispersal, competition, and landscape heterogeneity. *Oikos*, 116(7):1106–1119, 2007.

[70] A Goetschy and S E Skipetrov. Non-Hermitian Euclidean random matrix theory. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 84(1), 2011.

[71] Zhidong Bai and Jack W Silverstein. *Spectral analysis of large dimensional random matrices*. Springer, 2009.

[72] Per Bak. *How nature works: the science of self-organized criticality*, volume 383. Copernicus New York, 1996.

[73] Bertrand Russell. *History of Western Philosophy: Collectors Edition*. Routledge, 2013.

[74] John Toner, Yuhai Tu, and Sriram Ramaswamy. Hydrodynamics and phases of flocks. *Annals of Physics*, 318(1):170–244, July 2005.

[75] Jacopo Grilli, Sandro Azaele, Jayanth R. Banavar, and Amos Maritan. Absence of detailed balance in ecology. *EPL (Europhysics Letters)*, 100(3):38002, November 2012.

[76] Thierry Mora and William Bialek. Are Biological Systems Poised at Criticality? *Journal of Statistical Physics*, (May):268–302, June 2011.

[77] Gasper Tkacik, Elad Schneidman, Michael J Berry II, and William Bialek. Ising models for networks of real neurons. November 2006.

[78] Igor Volkov, Jayanth R Banavar, Stephen P Hubbell, and Amos Maritan. Inferring species interactions in tropical forests. *Proceedings of the National Academy of Sciences of the United States of America*, 106(33):13854–9, August 2009.

[79] Matteo Adorisio, Jacopo Grilli, Samir Suweis, Sandro Azaele, Jayanth R. Banavar, and Amos Maritan. Spatial maximum entropy modeling from presence/absence tropical forest data. July 2014.

[80] J M Beggs and D Plenz. Neuronal avalanches in neocortical circuits. *The Journal of neuroscience*, 23(35):11167–11177, 2003.

[81] D R Chialvo. Emergent complex neural dynamics. *Nature Physics*, 6:744–750, 2010.

[82] J M Beggs and N Timme. Being critical of criticality in the brain. *Frontiers in Physiology*, 3, 2012.

[83] Michele Ballerini, Nicola Cabibbo, Raphael Candelier, Andrea Cavagna, Evaristo Cisbani, Irene Giardina, Alberto Orlandi, Giorgio Parisi, Andrea Procaccini, Massimiliano Viale, and Vladimir Zdravkovic. Empirical investigation of starling flocks: a benchmark study in collective animal behaviour. *Animal Behaviour*, 76(1):201–215, July 2008.

[84] M Ballerini, N Cabibbo, R Candelier, A Cavagna, E Cisbani, I Giardina, V Lecomte, A Orlandi, G Parisi, A Procaccini, M Viale, and V Zdravkovic. Interaction ruling animal collective behavior depends on topological rather than metric distance: evidence from a field study. *Proceedings of the National Academy of Sciences of the United States of America*, 105(4):1232–7, January 2008.

[85] Andrea Cavagna, Alessio Cimarelli, Irene Giardina, Giorgio Parisi, Raffaele Santagati, Fabio Stefanini, and Massimiliano Viale. Scale-free correlations in starling flocks. *Proceedings of the National Academy of Sciences of the United States of America*, 107(26):11865–70, June 2010.

[86] Alessandro Attanasi, Andrea Cavagna, Lorenzo Del Castello, Irene Giardina, Tomas S Grigera, Asja Jelić, Stefania Melillo, Leonardo Parisi, Oliver Pohl, Edward Shen, and Massimiliano Viale. Information transfer and behavioural inertia in starling flocks. *Nature physics*, 10(9):615–698, September 2014.

[87] Alessandro Attanasi, Andrea Cavagna, Lorenzo Del Castello, Irene Giardina, Stefania Melillo, Leonardo Parisi, Oliver Pohl, Bruno Rossaro, Edward Shen, Edmondo Silvestri, and Massimiliano Viale. Collective behaviour without collective order in wild swarms of midges. *PLoS computational biology*, 10(7):e1003697, July 2014.

[88] William Bialek, Andrea Cavagna, Irene Giardina, Thierry Mora, Oliver Pohl, Edmondo Silvestri, Massimiliano Viale, and Aleksandra M Walczak. Social interactions dominate speed control in poising natural flocks near criticality. *Proceedings of the National Academy of Sciences of the United States of America*, 111(20):7212–7, May 2014.

[89] Alessandro Attanasi, Andrea Cavagna, Lorenzo Del Castello, Irene Giardina, Stefania Melillo, Leonardo Parisi, Oliver Pohl, Bruno Rossaro, Edward Shen, Edmondo Silvestri, and Others. Wild swarms of midges linger at the edge of an ordering phase transition. *arXiv preprint arXiv:1307.5631*, 2013.

[90] Tamás Vicsek, András Czirók, Eshel Ben-Jacob, Inon Cohen, and Ofer Shochet. Novel Type of Phase Transition in a System of Self-Driven Particles. *Physical Review Letters*, 75(6):1226–1229, August 1995.

[91] Marco Cosentino Lagomarsino, Alessandro L Sellerio, Philip D Heijning, and Bruno Bassetti. Universal features in the genome-level evolution of protein domains. *Genome biology*, 10(1):R12, January 2009.

[92] Jacopo Grilli, Bruno Bassetti, Sergei Maslov, and Marco Cosentino Lagomarsino. Joint scaling laws in functional and evolutionary categories in prokaryotic genomes. *Nucleic acids research*, 40(2):530–40, January 2012.

[93] X. Gabaix. Zipf's Law for Cities: An Explanation. *The Quarterly Journal of Economics*, 114(3):739–767, August 1999.

[94] K Okuyama, M Takayasu, and H Takayasu. Zipf's law in income distribution of companies. *Physica A: Statistical Mechanics and its Applications*, 269(1):125–131, July 1999.

[95] Daniel F. Merriam and John C. Davis. Using Zipf's Law to Predict Future Earthquakes in Kansas. *Transactions of the Kansas Academy of Science*, 112(1 & 2):127–129, March 2009.

[96] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality: An explanation of the 1/f noise. *Physical review letters*, 59(4):381–384, 1987.

[97] David J. Schwab, Ilya Nemenman, and Pankaj Mehta. Zipf's Law and Criticality in Multivariate Data without Fine-Tuning. *Physical Review Letters*, 113(6):068102, August 2014.

[98] W. Li. Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on Information Theory*, 38(6):1842–1845, 1992.

[99] Iacopo Mastromatteo and Matteo Marsili. On the criticality of inferred models. page 6, February 2011.

[100] William Blake. *A memorable fancy*. Apiary Press, 1965.

[101] C G Langton. Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D: Nonlinear Phenomena*, 42(1):12–37, 1990.

[102] N Bertschinger and T Natschlager. Real-time computation at the edge of chaos in recurrent neural networks. *Neural Computation*, 16(7):1413–1436, 2004.

[103] S A Kauffman. *The origins of order: Self-organization and selection in evolution*. Oxford University Press, USA, 1993.

[104] Woodrow L Shew and Dietmar Plenz. The Functional Benefits of Criticality in the Cortex. *The Neuroscientist*, 19(1):88–100, 2013.

[105] J M Beggs. The Criticality Hypothesis: How Local Cortical Networks Might Optimize Information Processing. *Phil. Trans. R. Soc. A*, 366(1864):329–343, 2008.

[106] Andreas Wagner. *Robustness and evolvability in living systems*. Princeton University Press Princeton, 2005.

[107] Claudius Gros. *Complex and adaptive dynamical systems: A primer*. Springer, 2008.

[108] Hidde De Jong. Modeling and simulation of genetic regulatory systems: a literature review. *Journal of computational biology*, 9(1):67–103, 2002.

[109] Sui Huang. Cell lineage determination in state space: a systems view brings flexibility to dogmatic canonical rules. *PLoS biology*, 8(5):e1000380, 2010.

[110] Matti Nykter, Nathan D Price, Maximino Aldana, Stephen A Ramsey, Stuart A Kauffman, Leroy E Hood, Olli Yli-Harja, and Ilya Shmulevich. Gene expression dynamics in the macrophage exhibit criticality. *Proceedings of the National Academy of Sciences*, 105(6):1897–1900, 2008.

[111] Dmitry Krotov, Julien O Dubuis, Thomas Gregor, and William Bialek. Morphogenesis at criticality. *Proceedings of the National Academy of Sciences*, 111(10):3683–3688, 2014.

[112] Enrique Balleza, Elena R Alvarez-Buylla, Alvaro Chaos, Stuart Kauffman, Ilya Shmulevich, and Maximino Aldana. Critical dynamics in genetic regulatory networks: examples from four kingdoms. *PLoS One*, 3(6):e2456, 2008.

[113] T M Cover and J Thomas. *Elements of Information Theory*. Wiley, 1991.

[114] Jorge Hidalgo, Jacopo Grilli, Samir Suweis, Miguel A Muñoz, Jayanth R Banavar, and Amos Maritan. Information-based fitness and the emergence of criticality in living systems. *Proceedings of the National Academy of Sciences*, 111(28):10095–10100, July 2014.

[115] Harry Eugene Stanley. *Introduction to phase transitions and critical phenomena*. Oxford University Press, 1971.

[116] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[117] I N Sanov. *On the probability of large deviations of random variables*, volume 42. 1957.

[118] Marc Mezard and Andrea Montanari. *Information, physics, and computation*. OUP Oxford, 2009.

[119] David H. Ackley, Geoffrey E. Hinton, and Terrence J. Sejnowski. A Learning Algorithm for Boltzmann Machines. *Cognitive Science*, 9(1):147–169, January 1985.

[120] David E Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Professional, 1989.

[121] Edo Kussell and Stanislas Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309(5743):2075–2078, 2005.

[122] Denise M Wolf, Vijay V Vazirani, and Adam P Arkin. Diversity in times of adversity: probabilistic strategies in microbial survival games. *Journal of theoretical biology*, 234(2):227–253, 2005.

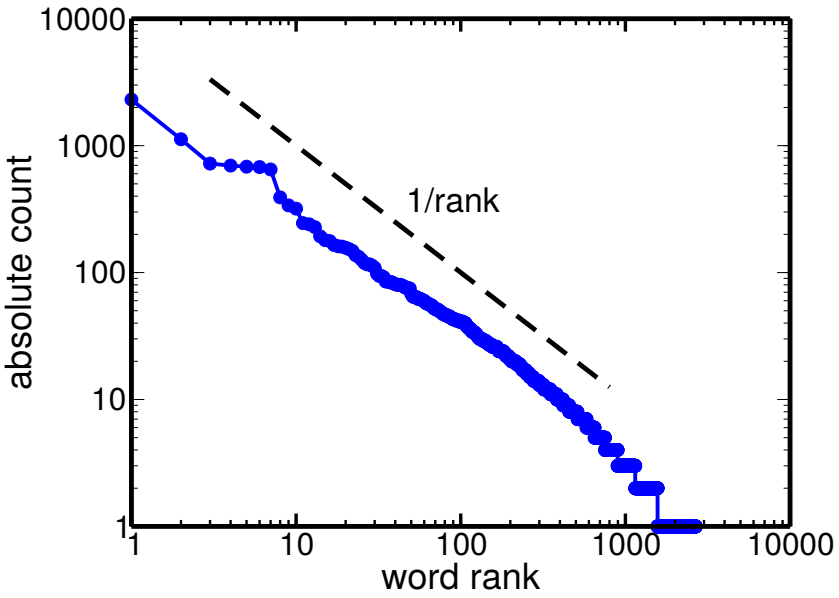
[123] A Goudarzi, C Teuscher, N Gulbahce, and T Rohlf. Emergent criticality through adaptive information processing in Boolean networks. *Phys Rev Lett*, 108(12):128702, 2012.

[124] Alan J McKane and T. J. Newman. Predator-Prey Cycles from Resonant Amplification of Demographic Stochasticity. *Physical Review Letters*, 94(21):1–4, June 2005.

[125] Crispin W Gardiner and Others. *Handbook of stochastic methods*, volume 3. Springer Berlin, 1985.

[126] N.G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. North Holland, 1981.

[127] Claude Itzykson and Jean-Michel Drouffe. *Statistical Field Theory: Volume 1, From Brownian Motion to Renormalization and Lattice Gauge Theory*. Cambridge University Press, 1991.



This thesis is (almost) Zipfian!
(thanks to Jordi Hidalgo for the idea)