Syntrophic interaction in microbial communities

Léo Buchenel

May 1, 2020

It's not what you get out of life that counts. Break your mirrors! In our society that is so self-absorbed, begin to look less at yourself and more at each other. You'll get more satisfaction from having improved your neighborhood, your town, your state, your country, and your fellow human beings than you'll ever get from your muscles, your figure, your automobile, your house, or your credit rating. You'll get more from being a peacemaker than a warrior.

R. Sargent Shriver, quoted by Arnold Schwarzenegger

Contents

1 Introduction

1.1 Consumer Resource Models in microbial ecology

Biology and Physics have always been tightly intertwined. Especially the years following the end of World War II saw many famous physicists getting interested in the blooming field of Biology [1], Leo Szilard or Erwin Schroedinger and his *What is Life? The Physical Aspect of the Living Cell* [2] among others. That exodus is no surprise, many biological phenomena at different scales are well modelled with Physics weaponry: from the use of Statistical Physics to solve protein folding problems [3] and find phase transitions in ecological communities [4] to the application of Hamiltonian dynamics to describe the movement of starling flocks [5].

However, Physics has not solved every problem yet: the study of microbial communities remains one of the biggest and most interesting challenges of contemporal microbiology. Indeed microbes and their complex interactions have a substantial, non trivial and very large impact on humans and their environment in various ways: we only start to understand the role of microbiological interactions in verterbrates' guts [6], or how they shape our soils [7] and oceans [8].

Population dynamics in ecological communities are often approximated by variations of the Lotka-Volterra model [9]. This approach works well when the mediators of the competitive interaction between species reach a steady state fast enough such that their own dynamics can be eliminated [10]. However, such an assumption is not always true and one must in general always ask themselves whether it may be applied [11]. For microbial communities, previous literature shows that the population dynamics are not always well captured by a Lotka-Volterra model [10], which explains the need of a more mechanistic approach, where the dynamics of both the microbes and their resources are explicitly modelled. Robert MacArthur is one of the first ecologists to establish and study such a *Consumer Resource Model* (CRM) [12], launching a field still active today [13].

In the light of recent developments in the microbiology literature [14], we propose here a CRM[1] which explicitly takes into account syntrophy. This process, which is largely observed in microbial communities [14], by definition occurs when microbes release, through a metabolic process, byproducts that are consumed by some members of the microbial community. In short, we want to know what happens when consumers are also allowed to release resources. The stability of an ecosystem, and how it is linked to its complexity, has always been of interest for ecologists [17]. **continuer ici**

1.2 Establishing the model and goals

We write down a CRM which describes the coupled evolution of the biomass of N_S different species and their N_R resources in a chemostat[2]. Resources are labelled $\mu = 1, \dots, N_R$

[1] One could argue that Flux Balance Analysis (FBA) [15] would be well suited for such a study. We ruled it out because it is known to scale badly [16] with system size and we do not want to be hindered by this limitation.

[2] In a chemostat, new nutrients are continuously added, while at the same time microorganisms and resources are removed in order to keep the culture volume constant [18].

and consumers $i = 1, \dots, N_S$. The coupled time evolution of their respective abundances $\{R_\mu, S_i\}$ is given by:

$$\boxed{\frac{dR_\mu}{dt} = l_\mu - m_\mu R_\mu - \sum_j \gamma_{j\mu} R_\mu S_j + \sum_j \alpha_{\mu j} S_j} \tag{1.1a}$$

$$\boxed{\frac{dS_i}{dt} = \sum_\nu \sigma_{i\nu}\gamma_{i\nu} R_\nu S_i - d_i S_i - \sum_\nu \alpha_{\nu i} S_i} \tag{1.1b}$$

The set of quantities $\{l_\mu, m_\mu, \gamma_{i\mu}, \alpha_{\mu i}, \sigma_{i\mu}, d_i\}$ has no explicit dynamics and is taken as constant. On the other hand, $\{R_\mu, S_i\}$ may dynamically evolve and will be refered to as *dynamical variables*. Note that there are in this model a lot of different symbols that may be easy to confuse. We will at least try to keep the following conventions:

- Quantities related to resources have subscripts in greek alphabet (*e.g.* the resource μ has abundance R_μ). Quantities related to species have subscripts in latin alphabet (*e.g.* the species i has abundance S_i). Finally, quantities related to both have both indices.
- Vectors (*i.e.* quantities with one index) are written with the latin alphabet (*e.g.* the resource μ has death rate m_μ).
- Matrices (*i.e.* quantities with two indices, usually relating resources and species) are written with the greek alphabet (*e.g.* $\gamma_{i\mu}$ is the rate at which species i consumes resource μ).

Our model takes numerous phenomena into account and it may be helpful to take the time to explain the different terms of each differential equation. The temporal evolution of the biomass R_μ of a resource μ is essentially driven by the following processes:

- Constant external inflow coming from the experimental setup: this corresponds to the constant $+l_\mu$ term.
- Natural diffusion/deterioration at rate m_μ: this corresponds to the $-m_\mu R_\mu$ term.
- Consumption by the species j at a rate $\gamma_{j\mu}$: $-\gamma_{j\mu} R_\mu S_j$. Summing up the contributions of every species, we get the Lotka-Volterra style [9] term $-\sum_j \gamma_{j\mu} R_\nu S_j$,
- Intrasystemic inflow coming from the syntrophy of species j at a rate $\alpha_{\mu j}$: $+\sum_j \alpha_{\mu j} S_j$.

On the other hand, biomass of species S_i changes because of the following processes:

- Consumption of resource R_ν at a rate $\gamma_{i\nu}$. Only a fraction $\sigma_{i\nu}$ of this is allocated to biomass growth: $+\sum_\nu \sigma_{i\nu}\gamma_{i\nu} R_\nu S_i$.
- Cell death/diffusion at rate d_i: this is the $-d_i S_i$ term.
- Syntrophic interaction : release of resource ν at rate $\alpha_{\nu i}$. In total $-\sum_\nu \alpha_{\nu i} S_i$.

The aim of the project is to study equilibria points of this model and their stability. In particular, we are interested in how syntrophy changes the robustness of the equilibria.

1.2.1 Equilibria of the model

We say that abundances $\{R^*_\mu, S^*_j\}$ are an *equilibrium*[3] of our model if they are fixed points of it, that means if the following equations are fulfilled:

$$\begin{cases} 0 = l_\mu - m_\mu R^*_\mu - \sum_j \gamma_{j\mu} R^*_\mu S^*_j + \sum_j \alpha_{\mu j} S^*_j & \text{(1.2a)} \\ 0 = \sum_\nu \sigma_{i\nu}\gamma_{i\nu} R^*_\nu S^*_i - d_i S^*_i - \sum_\nu \alpha_{\nu i} S^*_i & \text{(1.2b)} \end{cases}$$

1.2.2 Attack strategy and important notions

Before jumping right into the matter, it is important to explain how we will study this system of differential equations. Mainly two different but complimenteray approaches will be used: analytical and numerical. Note that the $\sim$ 5'000 lines of code we wrote from scratch and that we use to get the results of Section 3 are available at the address `https://gitlab.ethz.ch/palberto/consumersresources.git`.

Metaparameters and matrix properties

Studying the equilibria of our CRM will lead us to establish and study several relations involving the different *parameters* of the problem. Namely, these are: $l_\mu, m_\mu, \gamma_{i\mu}, \alpha_{\mu i}, \sigma_{i\mu}, d_i, R^*_\mu$ and S^*_i $\forall i = 1, \dots, N_S; \mu = 1, \dots, N_R$. We define the *parameters space* $\mathcal{P}$ as the space that contains all the parameters:

$$\mathcal{P} \equiv \left\{ p : p = (l_\mu, m_\mu, d_i, \gamma_{i\mu}, \alpha_{\mu i}, \sigma_{i\mu}, R^*_\mu, S^*_i) \right\} \tag{1.3}$$

Without taking into account the constraints on these parameters, there are $3N_R + 2N_S + 3N_R N_S$ free parameters, so $\mathcal{P} \simeq \mathbb{R}_+^{3N_R+2N_S+3N_R N_S}$. Our goal is to study microbial communities with a large number of consumers and resources, typically $N_R, N_S \approx 25, 50, 100, \dots$ *i.e.* $\mathcal{P} \simeq \mathbb{R}^{\sim 2000}$. It is clear that a precise study on each one of the 2000 elements is way too tenuous of a job. Another, simpler, approach is needed.

We decide to look at the problem from a statistical point of view, *i.e.* we write a matrix $q_{i\mu}$ as [19]:

$$q_{i\mu} = \mathfrak{Q} Q_{i\mu} \tag{1.4}$$

where $\mathfrak{Q}$ is a random variable of mean Q_0 and standard deviation σ_Q. $Q_{i\mu}$ is a binary matrix that, if interpreted as an adjacency matrix, tells about the network structure of the quantity $q_{i\mu}$.

[3] For the sake of brevity, we will sometimes drop the μ and j subscripts and simply write $\{R^*, S^*\}$.

We apply this way of thinking to the parameters of our problem, namely we write:

$$\begin{cases} l_\mu = \mathfrak{L} & \text{(1.5a)} \\ m_\mu = \mathfrak{M} & \text{(1.5b)} \\ \gamma_{i\mu} = \mathfrak{G} G_{i\mu} & \text{(1.5c)} \\ \alpha_{\mu i} = \mathfrak{A} A_{\mu i} & \text{(1.5d)} \\ \sigma_{i\mu} = \mathfrak{S} & \text{(1.5e)} \\ d_i = \mathfrak{D} & \text{(1.5f)} \\ R^*_\mu = \mathfrak{R} & \text{(1.5g)} \\ S^*_i = \mathfrak{S} & \text{(1.5h)} \end{cases}$$

Note that we do not add any explicit topological structure on $l_\mu, m_\mu, d_i, R^*_\mu, S^*_i$ and $\sigma_{i\mu}$ because we require these to always be larger than zero. In particular, we require positive-valued equilibria [20].

In order to make computations analytically tractable, we require the standard deviation on the parameters involved in the problem to be small, *i.e.* not larger than typically 10%. In that regime, every random variable $\mathcal{Q}$ is well approximated by its average value Q_0. We call Q_0 a *metaparameter*. While studying things analytically we will hence often come back to the following approximation:

$$\begin{cases} l_\mu \approx l_0 & \text{(1.6a)} \\ m_\mu \approx m_0 & \text{(1.6b)} \\ \gamma_{i\mu} \approx \gamma_0 G_{i\mu} & \text{(1.6c)} \\ \alpha_{\mu i} \approx \alpha_0 A_{\mu i} & \text{(1.6d)} \\ \sigma_{i\mu} \approx \sigma_0 & \text{(1.6e)} \\ d_i \approx d_0 & \text{(1.6f)} \\ R^*_\mu \approx R_0 & \text{(1.6g)} \\ S^*_i \approx S_0 & \text{(1.6h)} \end{cases}$$

This simplification is mathematically equivalent to collapsing the parameter space $\mathcal{P}$ to a lower dimensional space. Formally that lower dimensional space is the Cartesian product of $\mathcal{M}$ and $\mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S}$, where $\mathcal{M}$ is the *metaparameters space*:

$$\mathcal{M} \equiv \{m : m = (l_0, m_0, d_0, \gamma_0, \alpha_0, \sigma_0, R_0, S_0)\} \tag{1.7}$$

and $\mathcal{B}_{N \times M}$ is the set of binary matrices of dimensions $N \times M$. To summarize, we simply designed a *collapsing procedure* $\mathcal{C} : \mathcal{P} \to \mathcal{M} \times \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S}$ in order to simplify our problem.

Mathematically, when we do analytical computations, we mostly work in the collapsed space $\mathcal{C}(\mathcal{P})$ because it reduces the number of parameters from $3N_R + 2N_S + 3N_R N_S$ (continous) to 8 (continous) + $3N_R N_S$ (binary). And to make the problem even simpler, instead of looking at each entry of the binary matrices G and A individually, we will consider only some globally defined quantities of these matrices. For a matrix M_{ij} the metrics interesting to us are most of all:

- its **nestedness**[4]: this measures how "nested" the system is, *i.e.* if there are clusters grouped together[5]. It is known [19, 22] that nestedness can play a profound role in the dynamics of ecological communities. Although it is somewhat controversed [23], we will keep the definition of the nestedness $\eta(M)$ of a binary matrix M as it was used in [22]:

$$\eta(M) \equiv \frac{\sum_{i<j} n_{ij}}{\sum_{i<j} \min(n_i, n_j)} \tag{1.8}$$

  where the number of links n_i is simply the degree of the i-th row of M

$$n_i \equiv \sum_k M_{ik}, \tag{1.9}$$

  and n_{ij} is the overlap matrix defined as

$$n_{ij} \equiv \sum_k M_{ik} M_{jk}. \tag{1.10}$$

- its **connectance**: this measure, simply defined as the ratio of non-zero links in a matrix, is central in the study of plants-and-animals systems [19]. It is formally defined for a matrix q_{ij} of size $N \times M$ as:

$$\kappa(q) \equiv \frac{\sum_{ij} Q_{ij}}{NM} \tag{1.11}$$

  where Q is the (binary) network adjacency matrix of q.

Losing complexity – how to gain it back

As explained above, the idea is to simplify the study of a system with a large number of parameters to a system with a manageable number of so-called "metaparameters". Of course, collapsing a very high dimensional space to a low-dimensional space makes us lose information. Losing some information – and hence complexity – is desired when doing analytical computations but it is not when we want to produce precise and detailed numerical results.

So, how do we bridge the gap between what we work with analytically, *i.e.* a set of metaparameters and binary matrices, to precise measurements of quantities defined in our model Eq.(1.1)? The answer is simple: we define a function

$$\mathcal{A} : \mathcal{M} \times \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S} \to \mathcal{P} \tag{1.12}$$

which brings us from the collapsed space to the parameter space[6]. Numerically, from a set of metaparameters $m \in \mathcal{M}$ and binary matrices $B = (G, A) \in \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S}$, we produce a (or several) set(s) of parameters $p = \mathcal{A}(m, B) \in \mathcal{P}$ and study properties of it. Section 5.2.2 details how $\mathcal{A}$ is constructed.

[4]For the matrix consumption G, we will call it especially the "ecological overlap".

[5]In typical Lotka-Volterra models, where only species-species interactions are considered, *e.g.* [21], measuring the nestedness of the γ consumption matrix would be in the same spirit as counting how many niches there are in the community.

[6]Note that since the collapsed space is lower dimensional than the parameters space, $\mathcal{A}$ is not the inverse of $\mathcal{C}$.

2 Methods

2.1 Feasibility

Since its very inception [24], the study of ecological interactions has been and still is tightly close to the one of random matrices [25, 26, 27]. Usually, the procedure is we assume a feasible equilibrium point, where some matrix of the model (*e.g.* the species-interaction matrix or the jacobian) is approximated as random, and then study the dynamical or structural stability of said feasible point.

That framework is not satisfying for the study we would like to conduct, because it does not take time to study whether random parameters make sense in the first place. Indeed, before studying whether a microbial community can sustain perturbations, we need to know if said community actually *exists*. Biological systems, like any other natural systems, are constrained by laws, whether they arise from physical or biological considerations. For instance, it would not make sense to consider microbial communities that *e.g.* violate the laws of thermodynamics. In the following section, we explain how such considerations can help determining the answer to the *feasibility* question:

Can microbial communities arising from a random set of parameters make sense on a physical and biological level? If not, what are the conditions that should be imposed and how are these translated mathematically?

2.1.1 Basic concepts

As explained above, we want to impose conditions such that we only study systems that are compatible with biological and physical laws. Choosing such restrictions is a crucial task : we want to be as close to nature as possible but we also need to stay simple enough such that the model remains mathematically tractable. Our choice is the following : any system deemed as feasible must have "biological" model parameters and conserve biomass.

Asking for the model parameters to be "biological" means we want them to carry their intended biological interpretation. This means *e.g.* that any syntrophic interaction has to be non-negative $\alpha_{\mu i} \geqslant 0$ otherwise it cannot be interpreted as a syntrophic interaction anymore! More generally, the values of the parameters will be restricted. Namely, we are looking for positive-valued equilibria. Also, we require that every consumer can allocate some of each resource it consumes to growth[7] : zero efficiencies are forbidden. Finally every resource external feeding rate should be non-zero in order to avoid resource depletion and every resource and consumer must eventually die out in the absence of interaction. Mathematically, these considerations are equivalent to:

$$\boxed{R^*_\mu, S^*_i, \sigma_{i\mu}, l_\mu, d_i, m_\mu, \sigma_{i\mu} > 0 \text{ and } \gamma_{i\mu}, \alpha_{\mu i} \geqslant 0.} \tag{2.1}$$

That condition already greatly restricts the choice of parameters $p \in \mathcal{P}$. However, additional complexity arises from the relationships parameters have to follow by definition. Indeed, the $3N_R + 2N_S + 3N_RN_S$ parameters are constrained by the $N_R + N_S$ equations (1.2). So if we choose $2N_R + N_S + 3N_RN_S$ parameters, the remaining $N_R + N_S$ are instantly

[7] It wouldn't make sense to say that species i eats resource μ with efficiency 0, since this is equivalent to species i not eating resource μ, and this is already encoded in the network structure.

determined. Traditionally, we would solve for R^* and S^* and choose the rest of the parameters, but for reasons explained in Appendix 5.2.1, we will solve for the consumers death rates d_i and the resources diffusion rate m_μ. This means that if we *choose* non-negative $\gamma, \alpha, \sigma, \tau, l, R^*$ and S^*, Eqs.(2.1) can be combined with Eqs.1.2 into:

$$\begin{cases} d_i = \sum_\nu (\sigma_{i\nu}\gamma_{i\nu}R^*_\nu - \alpha_{\nu i}) > 0 \ \forall i = 1, \dots, N_S & \text{(2.2a)} \\ m_\mu = \dfrac{l_\mu - \sum_j \left(\gamma_{j\mu}R^*_\mu - \alpha_{\mu j}\right) S^*_j}{R^*_\mu} > 0 \ \forall \mu = 1, \dots, N_R & \text{(2.2b)} \end{cases}$$

In addition to these constraints, any feasible system should conserve biomass *at equilibrium*[8]: no species should be able to produce more biomass than it physically can. More specifically, a consumer i attains, from resources consumption, a total biomass of $\sum_\nu \gamma_{i\nu}R^*_\nu S^*_i$. From this available biomass, only $\sum_\nu \sigma_{i\nu}\gamma_{i\nu}R^*_\nu S^*_i$ is devoted to growth. Out of the remaining $\sum_\nu (1-\sigma_{i\nu})\gamma_{i\nu}R^*_\nu S^*_i$, a part $\sum_\nu \alpha_{\nu i}S^*_i$ is given back to the resources as a syntrophic interaction. We simply impose that the syntrophic interaction is smaller than or equal to the available remaining biomass :

$$\sum_\nu (1-\sigma_{i\nu})\gamma_{i\nu}R^*_\nu \geqslant \sum_\nu \alpha_{\nu i} \ \forall i = 1, \dots, N_S. \tag{2.3}$$

From now on, we will say that **a parameter set p is *feasible* if it satisfies Eqs.(2.2) and (2.3)**. This is completely deterministic, in the sense that for a given parameters set $p \in \mathcal{P}$ one can without a doubt say whether it is feasible or not. Hence we define the *parameters set feasibility function* $\mathfrak{F} : P \to \{0, 1\}$, which takes a parameter set as an input and tells you whether this parameter set is feasible or not:

$$\mathfrak{F}(p) = \begin{cases} 1 \text{ if } p \text{ is feasible,} \\ 0 \text{ else.} \end{cases} \tag{2.4}$$

However as explained in the introduction we will usually not work with a parameter set $p \in \mathcal{P}$ directly – because there are too many variables to keep track of – but with a metaparameter set $m \in \mathcal{M}$ and a consumption-syntrophy network $(G, A) \in \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S}$ instead. We can define a corresponding *metaparameters set feasibility function* $\mathcal{F} : \mathcal{M} \to [0, 1] \times \mathcal{B}_{N_R \times N_S}$ which is the probability that a given set of metaparameters $m \in \mathcal{M}$ coupled with binary matrices $B = (G, A)$ gives rise – through the algorithmic procedure $\mathcal{A}$ – to a feasible parameter set :

$$\boxed{\mathcal{F}(m, B) = \text{Probability}\left\{\mathfrak{F}\left(\mathcal{A}(m, B)\right) = 1\right\}} \tag{2.5}$$

In practice $\mathcal{F}(m, B)$ is estimated numerically by generating N parameters sets from (m, B) and calculating the number of feasible ones :

$$\mathcal{F}(m, B) = \lim_{N\to\infty} \sum_{i=1}^{N} \frac{\mathfrak{F}(\mathcal{A}(m, B))}{N} \approx \sum_{i=1}^{N} \frac{\mathfrak{F}(\mathcal{A}(m, B))}{N} \text{ for } N \gg 1. \tag{2.6}$$

[8]This weak condition should hold only at equilibrium : we allow transition periods where biomass may not be conserved.

2.1.2 The feasibility region

Appendix 5.2.2 explains the algorithmic procedure $\mathcal{A}(m, B) \in \mathcal{P}$ which allows us to build feasible parameters out of a set of metaparameters and a consumption-syntrophy network. However, in order to work properly, the combination of metaparameters used as an input of the algorithm must most of the time lead to the realisation of feasible systems. We hence need to find what region of the metaparameters space lead to a high feasibility : this is precisely the idea behind the notion of the feasibility region discussed below.

But first, let's see how our study can made simpler. Feasibility conditions discussed above tell us that we may choose six metaparameters[9] : γ_0, α_0, l_0, σ_0, S_0 and R_0. However, following the analysis of [27], we notice that our model Eqs.(1.1) still possesses some freedom. Indeed we can choose the set of units we work in to fix the values of some metaparameters. There are two physical quantities at stake here : biomass and time, and we may choose, however we want it, a specific set of units describing both of them. We measure biomass in units of the average resource abundance at equilibrium[10]:

$$\langle R_\mu \rangle = R_0 = 1. \tag{2.7}$$

Similarly, we measure time such that the average external resource uptake rate is one, that is:

$$\langle l_\mu \rangle = l_0 = 1. \tag{2.8}$$

After this manipulation, the number of metaparameters is reduced from six to four : only γ_0, S_0, α_0 and σ_0 remain.

For the sake of simplicity, we keep the same σ_0 throughout our whole study. We take a value close to the efficiency of real microbial systems [**insert ref**], that is $\sigma_0 = 0.25$.

Overall, we need to choose the last three remaining metaparameters: α_0, γ_0 and S_0. We will modify these metaparameters throughout the study. Since the remaining eight are fixed, we sometimes will elude them in the notation and will write instead of $m = (\gamma_0, S_0, \alpha_0, \sigma_0, R_0, l_0, d_0, m_0) \in \mathcal{M}$ simply $m = (\gamma_0, S_0, \alpha_0)$.

Formally, we can define for a consumption matrix G coupled with a syntrophy adjacency matrix A the x-feasible volume $\mathcal{F}_x^{G,A}$ of the metaparameters space $\mathcal{M}$ that will lead to at least a ratio x of feasible systems *i.e.* :

$$\mathcal{F}_x^{G,A} \equiv \{m \in \mathcal{M} : \mathcal{F}(m, (G, A)) \geqslant x\}. \tag{2.9}$$

It is clear that $\mathcal{F}_0^{G,A} = \mathcal{M}\ \forall G$ and $\text{Vol}\left(\mathcal{F}_x^{G,A}\right) \leqslant \text{Vol}\left(\mathcal{F}_y^{G,A}\right)\ \forall x > y, G$. We can similarly define for a set $S = \{(G_1, A_1), (G_2, A_2), \ldots, (G_N, A_N)\}$ of N couples of matrices their *common feasibility* region $\mathcal{F}_x^S$, which is the region of the metaparameters space where feasibility is at least x for every couple in the set:

$$\mathcal{F}_x^S \equiv \bigcap\nolimits_{(G,A)\in S} \mathcal{F}_x^{G,A}. \tag{2.10}$$

[9] Indeed, we saw that d_i and m_μ are set by the other parameters, so we cannot freely choose d_0 and m_0.

[10] That choice is not completely innocent. Indeed we will see later that the matrix $\alpha_{\nu i} - \gamma_{i\nu} R_\nu^*$ is a crucial quantity here. Setting $\langle R^* \rangle = 1$ allows us to simply study the impact of γ against α instead of the more complicated γR^* versus α.

We also define for a matrix set S, its critical feasibility $f^*(S)$, which is the largest feasibility we can get while still having a non-zero common volume :

$$f^*(S) \equiv \max_{x\in[0,1]} \left\{x : \text{Vol}\left(\mathcal{F}_x^S\right) > 0\right\}. \tag{2.11}$$

For actual computations, we will choose a matrix set S_M, stick to it during the whole thesis, and work in its critical feasibility region $\mathcal{F}^*$, defined as :

$$\boxed{\mathcal{F}^* \equiv \mathcal{F}^{S_M}_{f^*(S_M)}.} \tag{2.12}$$

Our hope is that we may find a fully feasible common region, *i.e.* $f^*(S) = 1$.

2.1.3 Estimating the fully feasible region $\mathcal{F}_1^{G,A}$

Now that we defined the x-feasible volume of a given couple consumption-syntrophy network (G, A) in Eq.(2.9), we would like to know what regions of the metaparameters space lead to fully feasible systems. We imposed two conditions that characterise the set of feasible parameters: Eqs.(2.2) and (2.3). We use them as a start to get corresponding metaparameters equations that describe $\mathcal{F}_1^{G,A}$.

Biomass conservation

As stated above, we require that biomass is conserved in our model. This is equivalent to fulfilling Eq.(2.3), which we rewrite here:

$$\sum_\nu (1-\sigma_{i\nu})\,\gamma_{i\nu} R^*_\nu \geqslant \sum_\nu \alpha_{\nu i} \;\forall i = 1,\ldots,N_S. \tag{2.13}$$

Eqs.(1.6) can be used to estimate the RHS of this equation:

$$\sum_\nu \alpha_{\nu i} \approx \deg(A,i)\alpha_0, \tag{2.14}$$

where $\deg(A,i)$ is the degree of the i-th column of the α matrix :

$$\deg(A,i) = \sum_\nu A_{\nu i}. \tag{2.15}$$

Similarly,

$$\sum_\nu (1-\sigma_{i\nu})\,\gamma_{i\nu} R^*_\nu \approx (1-\sigma_0) R_0 \sum_\nu \gamma_{i\nu} \approx \deg(G,i)(1-\sigma_0)R_0\gamma_0, \tag{2.16}$$

Energy conservation Eq.(2.3) is then equivalent to

$$\deg(A,i)\alpha_0 \lessapprox \deg(G,i)(1-\sigma_0)R_0\gamma_0 \;\forall i = 1,...,N_S \tag{2.17}$$

Since $\deg(G,i) > 0$, we have[11]:

$$\frac{\deg(A,i)}{\deg(G,i)}\alpha_0 \lessapprox (1-\sigma_0)R_0\gamma_0 \;\forall i = 1,...,N_S \tag{2.18}$$

[11]Indeed, $\deg(G,i)$ is the number of resources species i eats. We of course ask every consumer to at least consume something, otherwise they would not be part of the microbial community.

This is fulfilled if :

$$\boxed{\max_i \left\{ \frac{\deg(A,i)}{\deg(G,i)} \right\} \alpha_0 \lessapprox (1-\sigma_0) R_0 \gamma_0}. \tag{2.19}$$

Systems where the ratio $\frac{\#\text{resources released to}}{\#\text{resources consumed}}$ is small for each species allow for a larger individual syntrophy interaction (which is very intuitive).

Biological interpretation of the parameters

Additionally, the consumers death rates d_i have to be positive. This implied Eq.(2.2a), which may be recast as :

$$\sum_\mu \sigma_{i\mu} \gamma_{i\mu} R^*_\mu > \sum_\mu \alpha_{\mu i} \tag{2.20}$$

Using a reasoning similar to above, we get a corresponding metaparameters inequality:

$$\boxed{\max_i \left\{ \frac{\deg(A,i)}{\deg(G,i)} \right\} \alpha_0 \lessapprox \sigma_0 R_0 \gamma_0}. \tag{2.21}$$

Also, the resources diffusion rates m_ν need to be positive:

$$l_\nu + \sum_j \alpha_{\nu j} S^*_j > \sum_j \gamma_{j\nu} R^*_\nu S^*_j \ \forall \nu = 1, \ldots, N_R \tag{2.22}$$

Which is equivalent to

$$l_0 + \deg(A,\nu) \alpha_0 S_0 \gtrapprox \deg(G,\nu) \gamma_0 R_0 S_0 \ \forall \nu \tag{2.23}$$

Since $\deg(G,\nu) > 0$, we[12] can divide the above equations by $\deg(G,\nu) > 0$ and then recast these N_R equations into a single condition:

$$\boxed{\min_\nu \left\{ \frac{l_0}{\deg(G,\nu) S_0} + \frac{\deg(A,\nu)}{\deg(G,\nu)} \alpha_0 \right\} \gtrapprox \gamma_0 R_0} \tag{2.24}$$

This says that systems where the ratio $\frac{\#\text{number of species that release to me}}{\#\text{number of species that consume me}}$ is large for every resource are more feasible. The strategy should be then to have γ's that have large $\deg(G,\nu)$ (*i.e.* resources are consumed by many species) and large $\deg(G,i)$ (*i.e.* species consume a lot of species), and the other way around for α (not sure about this for the last one).

Combining both conditions

The two upper bounds Eqs.(2.19)-(2.21) on α_0 can be combined in a single inequality :

$$\max_i \left\{ \frac{\deg(A,i)}{\deg(G,i)} \right\} \alpha_0 \lessapprox \min(1-\sigma_0, \sigma_0) \gamma_0 R_0 \tag{2.25}$$

[12] Similarly to a previous footnote, we require that every resource ν is eaten by at least one consumer, *i.e.* $\deg(G,\nu) > 0$, otherwise it does not belong to the community.

Note that when $\alpha_0 > 0$, we will trivially require that the syntrophy matrix is not empty, *i.e.* there exists at least an i for which $\deg(A,i) \geqslant 1$. Also, the largest value $\deg(G,i)$ can get (for any i) is N_R. Hence,

$$\max_i \left\{ \frac{\deg(A,i)}{\deg(G,i)} \right\} \geqslant \frac{1}{N_R}, \tag{2.26}$$

and we can find the largest allowed theoretical non-zero α_0 :

$$\boxed{\alpha_0 \lessapprox \min(1-\sigma_0, \sigma_0)\gamma_0 R_0 N_R.} \tag{2.27}$$

Finally, Eq.(2.24) and (2.25) can be combined into a single one, which characterises the fully feasible region $\mathcal{F}_1^{G,A}$:

$$\boxed{\max_i \left\{ \frac{\deg(A,i)}{\deg(G,i)} \right\} \alpha_0 \lessapprox \min(1-\sigma_0, \sigma_0)\gamma_0 R_0 \lessapprox \min\left(1-\sigma_0, \sigma_0\right) \min_\nu \left\{ \frac{l_0}{\deg(G,\nu) S_0} + \frac{\deg(A,\nu)}{\deg(G,\nu)} \alpha_0 \right\}} \tag{2.28}$$

2.2 Dynamical stability

As stated in the introduction, our ultimate goal is to study equilibria points of the set of coupled differential equations (1.1). In particular we want to know how *stable* a given equilibrium is. However there is no consensual definition of stability: what does it mean exactly that a system is stable under a given perturbation? How is a perturbation even defined? Throughout this thesis different notions of stability will be tackled: the first is *dynamical stability*. The main idea behind dynamical stability is simple. We want to answer the following question:

Given an equilibrium point $\{R_\mu^*, S_i^*\}$, does the system go back to a positive-valued equilibrium when the consumers and resources abundances are changed? If yes, how much can they be changed before the system evolves in such a way that it does not reach a positive-valued equilibrium?

2.2.1 Definitions

Local dynamical stability

We first introduce *local dynamical stability*. A system is said to be *locally dynamically stable* if it goes back to *its initial equilibrium point* $\{R_\mu^*, S_i^*\}$ after R_μ^* and S_i^* have been perturbed by an infinitesimal amount $\{\Delta R_\mu(t_0), \Delta S_i(t_0)\}$ at time t_0.

More precisely, consider a system which is at equilibrium at time before $t = t_0$. Right after $t = t_0$, we perturb the equilibria abundances $\left\{R_\mu^*, S_i^*\right\}$ by an infinitesimal amount $\{\Delta R_\mu(t_0), \Delta S_i(t_0)\}$. We want to know how the perturbations away from equilibrium, written $\{\Delta R_\mu(t), \Delta S_i(t)\}$, and defined as

$$\Delta R_\mu(t) \equiv R_\mu(t) - R_\mu^* \text{ and } \Delta S_i(t) = S_i(t) - S_i^*. \tag{2.29}$$

will evolve qualitatively. Namely, will they go to zero or increase indefinitely as t increases? Perturbation analysis tells us **insert ref** that the quantity which drives the evolution of $\{\Delta R_\mu(t), \Delta S_i(t)\}$ is the *jacobian matrix of the system at equilibrium* J^*, given by :

$$J^* \equiv J(t_0), \tag{2.30}$$

where $J(t)$ is the *jacobian* of the system *i.e.* the jacobian matrix of its temporal evolution (1.1) evaluated at time t. $J(t)$ has a block matrix structure which is given by:

$$J(t) \equiv \begin{pmatrix} \frac{\partial \dot{R}_\mu}{\partial R_\nu} & \frac{\partial \dot{R}_\mu}{\partial S_j} \\ \frac{\partial \dot{S}_i}{\partial R_\nu} & \frac{\partial \dot{S}_i}{\partial S_j} \end{pmatrix} = \begin{pmatrix} \left(-m_\mu - \sum_j \gamma_{j\mu} S_j(t)\right) \delta_{\mu\nu} & -\gamma_{j\mu} R_\mu(t) + \alpha_{\mu j} \\ \sigma_{i\nu}\gamma_{i\nu} S_i(t) & \left(\sum_\nu \sigma_{i\nu}\gamma_{i\nu} R_\nu(t) - d_i - \sum_\nu \alpha_{\nu i}\right) \delta_{ij} \end{pmatrix}, \tag{2.31}$$

where δ is the ubiquitously occurring Kronecker delta symbol defined as:

$$\delta_{ij} = \begin{cases} 1 \text{ if } i = j, \\ 0 \text{ else.} \end{cases} \tag{2.32}$$

J^* is then precisely J with $\{R_\mu, S_i\}$ taken at the considered equilibrium point $\{R^*_\mu, S^*_i\}$, which simplifies its structure. Indeed, since we are interested only in positive valued equilibria (*i.e.* $S^*_i > 0\ \forall i$), then Eq.(1.2b) is equivalent to:

$$\sum_\nu \sigma_{i\nu}\gamma_{i\nu}R^*_\nu - d_i - \sum_\nu \alpha_{\nu i} = 0, \tag{2.33}$$

which means that the lower right block of the jacobian in Eq.(2.31) will be zero. Hence at equilibrium the jacobian J^* will have the following block form:

$$\boxed{J^* = \begin{pmatrix} -\Delta & \Gamma \\ \mathrm{B} & 0 \end{pmatrix}}, \tag{2.34}$$

where

- $\Delta_{\mu\nu} = \mathrm{diag}(m_\mu + \sum_j \gamma_{j\mu}S^*_j) = \mathrm{diag}\left(\frac{l_\mu + \sum_j \alpha_{\mu j}S^*_j}{R^*_\mu}\right)$ is a positive $N_R \times N_R$ diagonal matrix,
- $\Gamma_{\mu j} = -\gamma_{j\mu}R^*_\mu + \alpha_{\mu j}$ is a $N_R \times N_S$ matrix which does not have entries with a definite sign.
- $\mathrm{B}_{i\nu} = \sigma_{i\nu}\gamma_{i\nu}S^*_i$ is a $N_S \times N_R$ matrix with positive entries.

For reasons explained later in the manuscript, we say that a given equilibrium is *locally dynamically stable* if the largest real part of the eigenvalues of J^* is negative.

The locally dynamically stable volume $\mathcal{D}^{G,A}_{L,x}$

Similarly to what was conducted in Methods 2.1, one can define the *parameters set local dynamical stability function* $\mathfrak{D}_L : \mathcal{P} \to \{0, 1\}$, which tells you whether a given set of parameters $p \in \mathcal{P}$ is locally dynamically stable or not:

$$\mathfrak{D}_L(p) \equiv \begin{cases} 1 \text{ if } p \text{ is locally dynamically stable} \\ 0 \text{ else.} \end{cases} \tag{2.35}$$

Of course, p has to be feasible in order to be locally dynamically stable:

$$\mathfrak{D}_L(p) = 1 \implies \mathfrak{F}(p) = 1. \tag{2.36}$$

We also define the *metaparameters set local dynamical stability function* $\mathcal{D}_L : \mathcal{M} \times \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S} \to [0, 1]$ which tells you, given a set of metaparamters $m \in \mathcal{M}$ and a consumption-syntrophy network $B = (G, A)$ the chance that the procedure $\mathcal{A}(m, B)$ gives a locally dynamically stable set of parameters:

$$\boxed{\mathcal{D}_L(m, B) \equiv \text{Probability}\{\mathfrak{D}_L(\mathcal{A}(m, B)) = 1\}}. \tag{2.37}$$

We also define the x locally dynamically stable (lds) region $\mathcal{D}^{G,A}_{L,x}$ by the region of the metaparameters space that gives rise to a percentage of at least x dynamically stable systems:

$$\mathcal{D}^{G,A}_{L,x} \equiv \{m \in \mathcal{M} : \mathcal{D}_L(m, (G, A)) \geqslant x\} \tag{2.38}$$

Clearly, $\mathcal{D}_{L,0}^{G,A} = \mathcal{M}$, $\text{Vol}\left(\mathcal{D}_{L,x}^{G,A}\right) \leqslant \text{Vol}\left(\mathcal{D}_{L,y}^{G,A}\right)\ \forall x \geqslant y$, and more importantly, Eq.(2.36) is equivalent to $\mathcal{D}_{L,x}^{G,A} \subset \mathcal{F}_{x}^{G,A}$. We can also define for a set of N couples of matrices $S = \{(G_1, A_1) \ldots, (G_N, A_N)\}$ their common x lds-region $\mathcal{D}_{L,x}^{S}$:

$$\mathcal{D}_{L,x}^{S} \equiv \bigcap\nolimits_{(G,A)\in S} \mathcal{D}_{L,x}^{G,A}. \tag{2.39}$$

For such a set S we define also its critical local dynamical stability $d_L^*(S)$ which is the largest local dynamical stability we can achieve while still having a non-zero common volume:

$$d_L^*(S) = \max_{x\in[0,1]} \{x : \text{Vol}(\mathcal{D}_{L,x}) > 0\}\,. \tag{2.40}$$

Finally the critical common local dynamical stability volume $\mathcal{D}_L^S$ is the common lds-region at the critical local dynamical stability:

$$\mathcal{D}_L^* \equiv \mathcal{D}_{L,d_L^*(S)}^{S}. \tag{2.41}$$

The hope is that we can work most of the time with systems that have $d_L^*(S) = 1$.

Global dynamical stability

If we establish that a system is locally dynamically stable, we know that it will come back to the same equilibrium after an infinitesimal perturbation of the resources and consumers abundances. The next natural question is:

How much can these equilibria points be perturbed before the system goes to a point where either at least a species has gone extinct or reaches another positive valued equilibrium $\{\tilde{R}_\mu^*, \tilde{S}_i^*\}$ or simply does not reach a new dynamical equilibrium?

One way of studying this [19] is to simply take an equilibrium point $\{R_\mu^*, S_i^*\}$ and perturb the abundance of the species and resources at that point by a fixed number $\Delta_D \in [0, 1]$ which allows us to quantify the perturbation:

$$\begin{cases} R_\mu^* \to R_\mu(t_0) \equiv R_\mu^* \left(1 + \Delta_D \nu_\mu\right), & (2.42) \\ S_i^* \to S_\mu(t_0) \equiv S_i^* \left(1 + \Delta_D \nu_i\right), & (2.43) \end{cases}$$

where the $\nu_{\mu,i}$ are random numbers drawn from a uniform distribution between -1 and +1 and t_0 is the time where the previously at equilibrium system is perturbed. The system with the initial values $\{R(t_0), S(t_0)\}$ can then be time evolved from $t = t_0$ until it reaches an equilibrium $\{\tilde{R}^*, \tilde{S}^*\}$ which may be different from the equilibrium $\{R^*, S^*\}$ initially considered. This procedure is essentially a generalized version of local dynamical stability, since we allow the perturbation Δ_D to be non-infinitesimal. The question we will ask is precisely how big Δ_D can get.

A certain number of quantities, that all depend on the perturbation Δ_D, can then be measured to quantify the dynamical stability of the system:

- The resilience t_R: the time scale over which the system reaches its new equilibrium.

- The number of extinctions E: the number of species or resources which died during the time it took the system to reach its new equilibrium.
- The angle α between two equilibria: this quantifies how close the old and new equilibria are. α is defined through its standard scalar product formula:

$$\cos(\alpha) \equiv \frac{\sum_\mu R^*_\mu \tilde{R}^*_\mu + \sum_j S^*_j \tilde{S}^*_j}{\sqrt{\sum_\mu \left(R^*_\mu\right)^2 + \sum_i \left(S^*_i\right)^2}\sqrt{\sum_\mu \left(\tilde{R}^*_\mu\right)^2 + \sum_i \left(\tilde{S}^*_i\right)^2}}. \tag{2.44}$$

These quantities have either been already introduced in previous papers or are natural extensions of standard quantities [19, 28]. They allow us to quantify the robustness of a given equilibrium.

2.2.2 The quest for a full solution

The question of global dynamical stability is mathematically tedious, so we focus on local dynamical stability. We aim to find the spectrum of the jacobian at equilibrium, which will tell us whether the system is locally dynamically stable or not.

How to determine local dynamical stability

We stated above that the sign of the largest real part of all the eigenvalues of J^* determines the local dynamical stability. More precisely, we are interested in the real part of λ_1, which is defined by the following property:

$$\boxed{\forall \lambda \in \sigma(J^*),\ \mathrm{Re}\,(\lambda) \leqslant \mathrm{Re}\,(\lambda_1)}, \tag{2.45}$$

where $\sigma(J^*)$ is the set of eigenvalues of J^*, called the *spectrum* of J^*. Perturbation analysis tells us that the sign of the real part of λ_1 governs the local stability of the system at equilibrium **add source**. There are three cases:

- $\mathrm{Re}\,(\lambda_1) < 0$: any perturbation on the abundances is exponentially supressed. The system is stable.
- $\mathrm{Re}\,(\lambda_1) > 0$: any perturbation on the abundances is exponentially amplified. The system is unstable.
- $\mathrm{Re}\,(\lambda_1) = 0$: a second order perturbation analysis is required to assess the system's local dynamical stability. We call such systems *marginally stable* [29].

The master equation for local dynamical stability

In order to get $\mathrm{Re}\,(\lambda_1)$, we have to get the full spectrum of J^*, as a straight forward application of easier standard techniques like the Perron-Frobenius theorem [30] does not work. The eigenvalues of J^* are obtained through the eigenvalue problem:

$$\det\left(J^* - \lambda\right) = 0. \tag{2.46}$$

More explicitly, using Eq.(2.34), we state the *master equation for local dynamical stability*:

$$\boxed{\det \begin{pmatrix} -\Delta - \lambda & \Gamma \\ \mathsf{B} & 0 - \lambda \end{pmatrix} = 0} \tag{2.47}$$

That equation is not trivially solved. We then seek regimes where it could be made simpler.

Simplifying the master equation

Equation (2.47) may be simplified if

$$\lambda \neq 0. \tag{2.48}$$

Indeed[13], a non-zero λ implies

$$\det\left(\lambda \mathbb{1}_{N_S}\right) \neq 0, \tag{2.49}$$

where $\mathbb{1}_{N_S}$ stands for the $N_S \times N_S$ identity matrix. One can use this condition to simplify Eq.(2.47) using the properties of block matrices [31]:

$$\det \begin{pmatrix} -\Delta - \lambda \mathbb{1}_{N_R} & \Gamma \\ \mathsf{B} & 0 - \lambda \mathbb{1}_{N_S} \end{pmatrix} = \det\left(-\lambda \mathbb{1}_{N_S}\right) \det\left(-\Delta - \lambda \mathbb{1}_{N_R} + \frac{1}{\lambda}\Gamma\mathsf{B}\right). \tag{2.50}$$

Hence Eq.(2.47) becomes:

$$\boxed{\det\left(\lambda^2 \mathbb{1}_{N_R} + \Delta\lambda - \Gamma\mathsf{B}\right) = 0.} \tag{2.51}$$

The complexity here is already reduced because we go from the determinant of a $N_R + N_S$ square matrix to a N_R square matrix. We see from the previous expression that the dynamics is essentially dictated by the $\Gamma\mathsf{B}$ N_R-dimensional square matrix, which is given by:

$$\left(\Gamma\mathsf{B}\right)_{\mu\nu} = \sum_i \Gamma_{\mu i}\mathsf{B}_{i\nu} = \sum_i \left(\alpha_{\mu i} - \gamma_{i\mu} R^*_\mu\right) \sigma_{i\nu}\gamma_{i\nu} S^*_i. \tag{2.52}$$

There are many strategies here to find regimes of stability. One is the so-called "Reductio ad absurdum", which is explored later in Methods 2.2.4.

2.2.3 Bounds on the eigenvalues

Before studying Equation (2.51), we would like to know more about the spectrum of J^*. The most critical question is knowing *where* we expect the eigenvalues of J^* to lie on the complex plane.

Gerschgorin circle theorem

Gerschgorin circle theorem [32] states that every eigenvalue of a $N \times N$ square matrix A is located in one of the N discs D_i defined by:

$$D_i \equiv \left\{ z \in \mathbb{C} : |z - A_{ii}| \leqslant \sum_{j \neq i} |A_{ij}| \right\}. \tag{2.53}$$

[13] Appendix 5.2.3 elaborates on when that condition is fulfilled.

In a more mathematical language:

$$\sigma(A) \subset \bigcup_{i=1}^{N} D_i. \tag{2.54}$$

Intuitively, the circle theorem tells us that the eigenvalues of a matrix deviate from the diagonal elements by a value bounded by the sum of the off-diagonal elements. It is then easy to see that if all the discs D_i are located to the left of the imaginary axis (*i.e.* the discs contain only numbers with a negative real part), then the eigenvalues of A are all negative. This corresponds to the following lemma:

Lemma 1. *If a $N-$dimensional square matrix A verifies the equations:*

$$Re\,(A_{ii}) + \sum_{j \neq i} |A_{ij}| < 0, \forall\ i = 1, \ldots, N, \tag{2.55}$$

then Re $(\lambda) < 0\ \forall \lambda\ \in \sigma(A)$.

Proof. Let $\lambda \in \sigma(A)$. By the circle theorem, there exists $k \in \{1, \ldots, N\}$ such that :

$$|\lambda - A_{kk}| \leqslant \sum_{j \neq k} |A_{kj}|\,. \tag{2.56}$$

We now use the complex identity:

$$|\lambda - A_{kk}| \geqslant \mathrm{Re}\,(\lambda - A_{kk}) = \mathrm{Re}\,(\lambda) - \mathrm{Re}\,(A_{kk})\,. \tag{2.57}$$

Equation (2.55) implies:

$$\sum_{j \neq k} |A_{kj}| < -\mathrm{Re}\,(A_{kk})\,. \tag{2.58}$$

Combining the two previous inequalities yields:

$$\mathrm{Re}\,(\lambda) - \mathrm{Re}\,(A_{kk}) \leqslant |\lambda - A_{kk}| \leqslant \sum_{j \neq k} |A_{kj}| < -\mathrm{Re}\,(A_{kk})\,. \tag{2.59}$$

Comparing the RHS and LHS of this inequality yields:

$$\mathrm{Re}\,(\lambda) < 0. \tag{2.60}$$

□

The Gerschgorin circle theorem allows us to get a precious bound on the modulus of each eigenvalue and hence on the one that decides the dynamics of the system λ_1. Indeed we know that all eigenvalues of J^* will be located in one of the $N_R + N_S$ discs of J^*. These are the "resources" discs:

$$D_\mu^R \equiv \left\{ z \in \mathbb{C} : |z + \Delta_\mu| \leqslant \sum_j |\Gamma_{\mu j}| \right\}\ \forall \mu = 1, \ldots, N_R, \tag{2.61}$$

and the "consumers" discs:

$$D_i^C \equiv \left\{ z \in \mathbb{C} : |z| \leqslant \sum_\nu |B_{i\nu}| \right\} \ \forall i = 1, \ldots, N_S. \tag{2.62}$$

According to the circle theorem Eq.(2.54), all eigenvalues will be in the union of these circles, *i.e.* there exists $\forall \lambda \in \sigma\left(J^*\right)$ at least one μ^* or one i^* such that:

$$|\lambda| \leqslant \sum_\nu |B_{i^*\nu}| \tag{2.63}$$

or

$$|\lambda + \Delta_{\mu^*}| \leqslant \sum_j |\Gamma_{\mu^* j}| \tag{2.64}$$

The triangle inequality implies:

$$|\lambda| \leqslant |\lambda + \Delta_{\mu^*}| + |-\Delta_{\mu^*}| \leqslant \sum_j |\Gamma_{\mu^* j}| + |-\Delta_{\mu^*}| = \sum_j |\Gamma_{\mu^* j}| + \Delta_{\mu^*}. \tag{2.65}$$

The only way both Eq.(2.63) and (2.65) are satisfied for all eigenvalues, and especially the one with the highest real part λ_1 is if they are bound by the maximum of both RHS of these equations. More precisely:

$$\boxed{|\lambda| \leqslant R_C \ \forall \lambda \in \sigma(J^*),} \tag{2.66}$$

where we defined the critical radius R_C as:

$$R_C \equiv \max \left\{ \max_i \left\{ \sum_\nu |B_{i\nu}| \right\}, \max_\mu \left\{ \sum_j |\Gamma_{\mu j}| + \Delta_\mu \right\} \right\}. \tag{2.67}$$

This gives us an estimation of how big the eigenvalues can get: we know that all the eigenvalues *have* an absolute value smaller than or equal to the critical radius R_C. The next step is to estimate R_C in terms of metaparameters, so that we can get a qualitative insight on how the eigenvalues change when the metaparameters are changed.
Using techniques very similar to previous computations, we estimate:

$$\sum_j |\Gamma_{\mu j}| + \Delta_\mu = \sum_j \left| \alpha_{\mu j} - \gamma_{j\mu} R_\mu^* \right| + \frac{l_\mu + \sum_j \alpha_{\mu j} S_j^*}{R_\mu^*} \approx \deg(\Gamma, \mu) \left| \alpha_0 - \gamma_0 R_0 \right| + \frac{l_0 + \deg(A, \mu) \alpha_0 S_0}{R_0}. \tag{2.68}$$

It is difficult to simplify $\deg(\Gamma, \mu) \approx \deg(A - G^T, \mu)$. If we assume that A and G have a low connectance then $\deg(A, \mu), \deg(G^T, \mu) \ll N_S$ and we may use the very loose approximation

$$\deg(A - G^T, \mu) \approx \deg(A, \mu) + \deg(G, \mu). \tag{2.69}$$

In that regime we then have:

$$\max_\mu \left\{ \sum_j |\Gamma_{\mu j}| + \Delta_\mu \right\} \approx \max_\mu \left\{ \left(\deg(A, \mu) + \deg(G, \mu) \right) \left| \alpha_0 - \gamma_0 R_0 \right| + \frac{l_0 + \deg(A, \mu) \alpha_0 S_0}{R_0} \right\}. \tag{2.70}$$

Similarly we find

$$\sum_{\nu} |B_{i\nu}| \approx \deg(G, i)\sigma_0\gamma_0 S_0, \tag{2.71}$$

such that R_C can be estimated roughly as:

$$R_C \approx \max\left\{\max_i \left(\deg(G,i)\right)\sigma_0\gamma_0 S_0\ , \right. \\ \left. \max_\mu \left\{ \left(\deg(A,\mu) + \deg(G,\mu)\right)|\alpha_0 - \gamma_0 R_0| + \frac{l_0 + \deg(A,\mu)\alpha_0 S_0}{R_0} \right\} \right\} \tag{2.72}$$

2.2.4 Low intra resources interaction (LRI) regime

Now that we have a bound on how big the eigenvalues can be, we need strategies to find regimes where we *know* $\mathrm{Re}\,(\lambda_1) < 0$, *i.e.* local dynamical stability is guaranteed. We inspire ourselves from the general idea of the mathematical proofs of [20].

Reductio ad absurdum

Using standard properties of determinants, we rewrite Eq.(2.51) as[14]:

$$\det\left(-\Delta^{-1}\right)\det\left(-\Delta^{-1}\lambda^2 - \lambda + \Delta^{-1}\Gamma\mathrm{B}\right) = 0 \iff \boxed{\det\left(S(\lambda) - \lambda\right) = 0} \tag{2.73}$$

with

$$\boxed{S(\lambda) = \Delta^{-1}\Gamma\mathrm{B} - \Delta^{-1}\lambda^2}, \tag{2.74}$$

or, component-wise:

$$S_{\mu\nu} = \frac{1}{\Delta_\mu}\left[\left(\sum_i \Gamma_{\mu i}\mathrm{B}_{i\nu}\right) - \lambda^2\delta_{\mu\nu}\right] \tag{2.75}$$

The strategy is to assume we are in an unstable regime,*i.e.* there exists at least one $\lambda \in \sigma(J^*)$ with $\mathrm{Re}\,(\lambda) \geqslant 0$ that satisfies Eq.(2.51) and such that $\mathrm{Re}\,(\lambda) > 0$. By Eq.(2.73), λ is also an eigenvalue of $S(\lambda)$. If we find conditions under which the real part of the spectrum of $S(\lambda)$ is entirely negative, we will know that $\mathrm{Re}\,(\lambda) \leqslant 0$. As this is a contradiction to the hypothesis that the regime is unstable, we must conclude that the regime is stable[15].

Hence, the general idea is to find regimes where we know that the spectrum of S, written as $\sigma(S)$, will be entirely negative for a positive λ. Thanks to the help of the two following theorems, we found such a regime, called *low intra-resource interaction* or LRI.

Strong LRI regime

Theorem 1. *Let p be a parameter set with a jacobian at equilibrium J^*. If 0 is not an eigenvalue of J^* and the equations*

$$(\Gamma B)_{\mu\mu} < -\sum_{\nu\neq\mu}\left|(\Gamma B)_{\mu\nu}\right| - R_C^2\ \forall\mu, \tag{2.76}$$

are verified, then p is dynamically stable.

[14]We can do this because since $m_\mu > 0$, we know Δ will always be invertible.

[15]Indeed, Eq.(2.51) assumes already that either $\mathrm{Re}\,(\lambda_1) > 0$ or $\mathrm{Re}\,(\lambda_1) < 0$.

Proof. We assume

$$(\Gamma \mathrm{B})_{\mu\mu} < -\sum_{\nu\neq\mu} \left|(\Gamma \mathrm{B})_{\mu\nu}\right| - R_C^2 \ \forall \mu. \tag{2.77}$$

This implies:

$$(\Gamma \mathrm{B})_{\mu\mu} + R_C^2 < -\sum_{\nu\neq\mu} \left|(\Gamma \mathrm{B})_{\mu\nu}\right| \ \forall \mu. \tag{2.78}$$

Using Eq.(2.66) and $\mathrm{Im}\,(\lambda)^2 \leqslant |\lambda|^2$, we get:

$$(\Gamma \mathrm{B})_{\mu\mu} + \mathrm{Im}\,(\lambda)^2 < -\sum_{\nu\neq\mu} \left|(\Gamma \mathrm{B})_{\mu\nu}\right| \ \forall \mu. \tag{2.79}$$

It is not difficult to prove that for any complex number:

$$\mathrm{Im}\,(c)^2 \geqslant -\mathrm{Re}\left(c^2\right) \ \forall c \in \mathbb{C}. \tag{2.80}$$

Using this result and dividing Eq.(2.79) by[16] Δ_μ, we get:

$$\frac{1}{\Delta_\mu}\left[\left(\sum_i \Gamma_{\mu i}\mathrm{B}_{i\mu}\right) - \mathrm{Re}\left(\lambda^2\right)\right] < -\sum_{\nu\neq\mu}\left|\frac{\sum_i \Gamma_{\mu i}\mathrm{B}_{i\nu}}{\Delta_\mu}\right| \ \forall \mu. \tag{2.81}$$

Looking at Eq.(2.75), we see that Eq.(2.81) is equivalent to:

$$\mathrm{Re}\,(S_{\mu\mu}) + \sum_{\nu\neq\mu} |S_{\mu\nu}| < 0 \ \forall \mu. \tag{2.82}$$

Lemma 1 implies that all the eigenvalues of $S(\lambda)$ have a negative real part. As explained before that means that if $\mathrm{Re}\,(\lambda_1) \geqslant 0$ in Eq.(2.74) (unstable or marginally stable regime), then $\mathrm{Re}\,(\lambda_1) < 0$, which leads to a contradiction. This then implies that the equilibrium is dynamically stable. □

Theorem 1 is a strong statement but it asks a lot on the parameters set, namely $\Gamma\mathrm{B}$ must have diagonal elements that are "very negative", which imposes severe conditions especially on the α and γ matrices. That is why we do not expect to find many parameters sets which would be in such a regime. We need to find another more relaxed regime in which more parameters sets could be.

Weak LRI regime

Another version of Theorem 1 can be stated. Its proof is in Appendix 5.2.4. Its assumptions are less restrictive, but that comes with a price : its statement is weaker.

Theorem 2. *Let p be a parameter set with a jacobian at equilibrium J^*. If 0 is not an eigenvalue of J^* and the equations*

$$(\Gamma B)_{\mu\mu} < -\sum_{\nu\neq\mu} \left|(\Gamma B)_{\mu\nu}\right| \ \forall \mu, \tag{2.83}$$

are verified, then the real eigenvalues of J^* are negative.

[16] That step is valid because $\Delta_\mu > 0$, $\forall \mu$.

Intuitive interpretation of the LRI regimes

Feasibility of the LRI regimes

So we found that if a system has parameters that respect Eq.(2.76) then it is dynamically stable. A naturally arising question is then to ask in what measure this is compatible with the feasibility equations Eqs.(2.2) and (2.3). The path is simple : we need to find a meta-parameters approximation of the resource interaction matrix $(\Gamma\mathrm{B})_{\mu\nu}$, which will allow us to give a "metaparameters-version" of Eq.(2.76) and compare it with the metaparameters feasibility equation (2.28).

Using the metaparameters approximations Eqs.(1.6), Eq.(2.52) can be simplified as:

$$(\Gamma\mathrm{B})_{\mu\nu} \approx \sigma_0\gamma_0 S_0 \left(\alpha_0 \sum_i A_{\mu i}G_{i\nu} - \gamma_0 R_0 \sum_i G_{i\mu}G_{i\nu}\right) \equiv \sigma_0\gamma_0 S_0 \left(\alpha_0 O_{\mu\nu} - \gamma_0 R_0 C_{\mu\nu}\right) \quad (2.84)$$

where we defined the *syntrophy overlap matrix* $O_{\mu\nu}$ and the *consumption overlap matrix* $C_{\mu\nu}$ as:

$$O_{\mu\nu} \equiv (AG)_{\mu\nu} \text{ and } C_{\mu\nu} \equiv \left(G^T G\right)_{\mu\nu}. \quad (2.85)$$

The clash syntrophy versus consumption between these two binary matrices essentially builds the dynamics of our model and an intuitive understanding of them can be very helpful.

The syntrophy overlap matrix $O_{\mu\nu}$ is defined as:

$$O_{\mu\nu} \equiv \sum_k A_{\mu k}G_{k\nu}. \quad (2.86)$$

Although A and G are binary, O does not have to and usually will not be. A given consumer k contributes to $O_{\mu\nu}$ if and only if both $A_{\mu k}$ and $G_{k\nu}$ are non zero, that is if consumer k releases resource μ *and* consumes resource ν. Hence $O_{\mu\nu}$ essentially tells how many species effectively "link" resource μ to resource ν through the indirect interaction of the species consumption.

Similarly, the consumption overlap matrix is defined as:

$$C_{\mu\nu} = \sum_k G_{k\mu}G_{k\nu}. \quad (2.87)$$

Like O, C usually is not binary. The intuition behind $C_{\mu\nu}$ is straight forward: it counts how many species eat both resource μ and ν.

We then find a lowerbound for the RHS of Eq.(2.76):

$$-\sum_{\nu\neq\mu} |\Gamma\mathrm{B}|_{\mu\nu} \geqslant -\sum_{\nu\neq\mu} \max_{\nu\neq\mu} |\Gamma\mathrm{B}|_{\mu\nu} \geqslant -\deg(\mu, O-C) \max_{\nu\neq\mu} |\Gamma\mathrm{B}|_{\mu\nu}. \quad (2.88)$$

Combining this with the approximation of $\Gamma\mathrm{B}$ above we get an approximative LRI regime condition on the metaparameters:

$$\boxed{\alpha_0 O_{\mu\mu} - \gamma_0 R_0 C_{\mu\mu} \lessapprox -\deg(\mu, O-C) \max_{\nu\neq\mu} |\alpha_0 O_{\mu\nu} - \gamma_0 R_0 C_{\mu\nu}| - \frac{R_C^2}{\sigma_0\gamma_0 S_0} \ \forall\mu.} \quad (2.89)$$

Since R_C gets smaller as the largest degree of G decreases (see Eq.2.72) we only expect systems with a low connectance food consumption adjacency matrix to be able to achieve an LRI state.

This allows to give a necessary condition on the magnitude of α_0. Indeed, since the RHS of the previous equation is negative, we need:

$$\alpha_0 O_{\mu\mu} - \gamma_0 R_0 C_{\mu\mu} < 0 \ \forall \mu \implies \alpha_0 \max_\mu \left\{ \frac{O_{\mu\mu}}{C_{\mu\mu}} \right\} < \gamma_0 R_0. \tag{2.90}$$

Systems where the maximal ratio of $O_{\mu\mu}$ and $C_{\mu\mu}$ is small will more likely lead to a LRI regime than others. The most favoured systems will be those where $O_{\mu\mu} = 0 \ \forall \mu$, *i.e.* systems where no species consumes what it itself produces. In that way we may say that coprophagy tends to destabilize microbial communities.

Combining Eq.(2.24) and (2.90) gives us a necessary condition on α_0 for feasible systems:

$$\boxed{\alpha_0 \left[\max_\mu \{O_{\mu\mu}\} - \min_\mu \left\{ \frac{\deg(A,\mu)}{\deg(G,\mu)} \right\} \right] \leqslant \frac{l_0}{\min_\mu \left(\deg(G,\mu) \right) S_0}}. \tag{2.91}$$

Although that equation gives us a necessary condition, it is not sufficient. Eq.(2.89), on the other hand, is and provides an intuitive way of finding a syntrophy adjacency matrix $A_{\mu i}$ that would put a system with a given consumption adjacency $G_{\mu i}$ in an LRI regime. The following section explains how this can be tried numerically.

Monte Carlo algorithm for the optimal syntrophy matrix

We want to find a general algorithm which, for a given food consumption adjacency matrix G gives back an optimal syntrophy adjacency matrix A. Strategically, we would like an A such that Eq.(2.89) is as close to being satisfied as possible. If it were satisfied, it would put the system in an LRI regime, which we have proven is dynamically stable.

Given a consumption network G and fixed metaparameters, Eq.(2.89) is more likely satisfied if the syntrophy matrix A is such that the LHS of Eq.(2.89) is minimal and its RHS is maximal. The LHS is minimized if $O_{\mu\mu} = (AG)_{\mu\mu}$ is set to its lowest possible value for every μ, that is zero. On the other hand, the RHS is maximal if $\alpha_0 (AG)_{\mu\nu} \approx \gamma_0 R_0 (G^T G)_{\mu\nu} \ \forall \nu \neq \mu$.

Intuitively, we then search for systems where AG is zero on the diagonal, *i.e.* where no coprophagy is observed, and $AG \approx \frac{\gamma_0 R_0}{\alpha_0} G^T G$ outside the diagonal. It can be formalized by writing a proper Metropolis-Hastings Markov Chain Monte Carlo (MCMC) method. We designed the following algorithmic procedure to build a syntrophy adjacency matrix A:

1. Create a random binary A. Its connectance is chosen as the one of the consumption matrix G.

2. Do the following for a given number of steps:

   - Choose a random row or, every other iteration, a column.

- In that row/column, try to swap a zero and a one while preserving the "releasers": if a species releases some resource, it has to keep releasing something (the resource can change though). The "releasees" are preserved as well: if a resource is being released by some species, it has to keep being released (but it does not have to be by the same species).
- The swap is accepted, *i.e.* A is modified, if the energy difference ΔE is negative or if a random number drawn uniformly between zero and one is smaller than $e^{-\Delta E/T}$ where T is the current temperature. More on ΔE and T below.

3. Return A.

A couple comments on this algorithm can be made:

- The algorithm preserves the connectance of A but not its nestedness. The question of what connectance to choose is open, but we choose $\kappa(A) = \kappa(G)$ as a first approach, *i.e.* syntrophy and consumption networks have the same connectance.

- The temperature T changes dynamically during the simulation. It is obtained in a way close to the spirit of simulated annealing techniques [33]: the temperature T is multiplied by a factor $\lambda = 0.99$ at a fixed frequency (for instance every 1000 steps). We add the requirement that if new moves are rejected during too many consecutive steps, we multiply the temperature by $1/\lambda$.

- The energy difference ΔE between the new proposed A' and the old A is computed by assigning an energy E to both A' and A and subtracting them:

$$\Delta E \equiv E(A', G) - E(A, G). \tag{2.92}$$

The choice of the energy function E is crucial. In essence, this MCMC algorithm will find the specific A which minimizes $E(A, G)$. Since we want to build systems in the LRI regime, we use the simplest and most natural function that is compatible with the intuitively expected characteristics of A explained above (*i.e.* AG is zero on the diagonal and equal to $\frac{\gamma_0 R_0}{\alpha_0} G^T G$ outside of it):

$$E(A, G) \equiv \sum_{\mu} \left(|\alpha_0 (AG)_{\mu\mu}| + \sum_{\nu \neq \mu} \left| (\alpha_0 AG - \gamma_0 R_0 G^T G)_{\mu\nu} \right| \right). \tag{2.93}$$

The energy function and hence the optimal syntrophy adjacency matrix A depend on the ratio $\frac{\alpha_0}{\gamma_0 R_0}$. This prompts then the question of which α_0 can be deemed sensible. As a first step, we will take the value of Eq.(2.27): $\alpha_0 = \min(1 - \sigma_0, \sigma_0)\gamma_0 R_0 N_R$. This means that the outcome of the algorithm is an optimized A *for the largest feasible syntrophy*. Since the expression we have for the largest feasible syntrophy is independent of the G matrix, this choice of α_0 provides us a sensible way of comparing different consumption networks.

2.2.5 Application: fully connected consumption and syntrophy networks

As an application, we consider the very special case of the fully connected consumption and syntrophy networks. In this "mean-field" theory, every consumer consumes and releases each resource, *i.e.*

$$G_{i\mu} = A_{\mu i} = 1 \ \forall \mu = 1, \dots, N_R \ \ \forall i = 1, \dots, N_S. \tag{2.94}$$

Our goal is to find the spectrum of systems with such consumption and syntrophy networks. Barbier and Arnoldi showed that the variance of the interaction matrix plays a leading role in the dynamics of their model [27]. We follow an approach similar to theirs and perform a *standard deviation expansion*.

Standard deviation expansion The idea behind the standard deviation expansion is the following. Let $q_{i\mu}$ be an arbitrary matrix of size $N_S \times N_R$. The nice trick done in [27] is to write the elements of the q matrix in terms of new variables $\tilde{q}_{i\mu}$:

$$q_{i\mu} = \langle q \rangle + \sigma_q \tilde{q}_{i\mu}. \tag{2.95}$$

In that expression, $\langle q \rangle$ is the average of q, element-wise

$$\langle q \rangle \equiv \frac{\sum_{\mu,i} q_{i\mu}}{N_S N_R}, \tag{2.96}$$

and σ_q is the standard deviation of q, again element-wise:

$$\sigma_q \equiv \sqrt{\langle q^2 \rangle - \langle q \rangle^2} \text{ with } \langle q^2 \rangle \equiv \frac{\sum_{\mu,i} q_{i\mu}^2}{N_S N_R}. \tag{2.97}$$

The main advantage of this procedure is that we get a clear idea about the scales involved. A matrix element $q_{i\mu}$ is roughly the mean $\langle q \rangle$ plus a deviation σ_q multiplied by a factor of magnitude ~ 1. Indeed the $\tilde{q}_{i\mu}$ are not large since they follow the two equalities [27]:

$$\langle \tilde{q} \rangle = 0 \text{ and } \langle \tilde{q}^2 \rangle = 1. \tag{2.98}$$

We apply this framework to our problem by noticing that if the $q_{i\mu}$ are all random samples coming from the same distribution law $\mathfrak{Q}$, we can write the following approximation in the case $N_R, N_S \gg 1$:

$$\langle q \rangle \approx \langle \mathfrak{Q} \rangle \equiv q_0. \tag{2.99}$$

We can then rewrite the free parameters of our model[17]:

$$\begin{cases} l_\nu \approx l_0 + \sigma_l \tilde{l}_\nu & \text{(2.100a)} \\ R_\nu^* \approx R_0 + \sigma_R \tilde{r}_\nu & \text{(2.100b)} \\ S_i^* \approx S_0 + \sigma_S \tilde{s}_i & \text{(2.100c)} \\ \gamma_{i\nu} \approx \gamma_0 + \sigma_\gamma \tilde{g}_{i\nu} & \text{(2.100d)} \\ \alpha_{\nu i} \approx \alpha_0 + \sigma_\alpha \tilde{\alpha}_{\nu i} & \text{(2.100e)} \\ \sigma_{i\nu} \approx \sigma_0 + \sigma_\sigma \tilde{\sigma}_{i\nu} & \text{(2.100f)} \end{cases}$$

[17] This works with γ and α because G and A have a trivial topology. Otherwise we would have to take their structure into account and the computations would not be as easy.

The strategy is to assume that the standard deviations are small which allows us to proceed to a first order Taylor expansion.

Rewriting the jacobian at equilibrium The different blocks of the jacobian at equilibrium (2.34) can be written with the new variables :

$$\begin{cases} \dfrac{l_\mu + \sum_j \alpha_{\mu j} S_j^*}{R_\mu^*} = \dfrac{l_0 + N_S \alpha_0 S_0 + \sigma_l \tilde{l}_\mu + \sigma_\alpha S_0 \sum_j \tilde{\alpha}_{\mu j} + \sigma_\alpha \sigma_S \sum_j \tilde{\alpha}_{\mu j} \tilde{s}_j}{R_0 + \sigma_R \tilde{r}_\mu} & (2.101) \\ -\gamma_{j\mu} R_\mu^* + \alpha_{\mu j} = -\gamma_0 R_0 + \alpha_0 + \sigma_\gamma R_0 \tilde{\gamma}_{j\mu} + \sigma_R \gamma_0 \tilde{r}_\mu + \sigma_\alpha \tilde{\alpha}_{\mu j} + \sigma_\gamma \sigma_R \tilde{\gamma}_{j\mu} \tilde{r}_\mu & (2.102) \\ \sigma_{i\nu} \gamma_{i\nu} S_i^* = \sigma_0 \gamma_0 S_0 + \sigma_\sigma \gamma_0 S_0 \tilde{\sigma}_{i\nu} + \sigma_\gamma \sigma_0 S_0 \tilde{\gamma}_{i\nu} + \sigma_S \sigma_0 \gamma_0 \tilde{s}_i + \sigma_\sigma \sigma_\gamma S_0 \tilde{\sigma}_{i\nu} \tilde{\gamma}_{i\nu} & \\ \quad + \sigma_\sigma \sigma_S \gamma_0 \tilde{\sigma}_{i\nu} \tilde{s}_i + \sigma_\gamma \sigma_S \sigma_0 \tilde{\gamma}_{i\nu} \tilde{s}_i + \sigma_\sigma \sigma_\gamma \sigma_S \tilde{\sigma}_{i\nu} \tilde{\gamma}_{i\nu} \tilde{s}_i & (2.103) \end{cases}$$

It's easier to work with relative standard deviations, *i.e.* we rewrite for all parameters :

$$\sigma_P \equiv \epsilon_P \langle P \rangle, \ \forall P \in \{l_\nu, R_\nu^*, S_i^*, \gamma_{i\nu}, \alpha_{\nu i}, \sigma_{i\nu}\}. \tag{2.104}$$

The previous relations then become :

$$\begin{cases} \dfrac{l_\mu + \sum_j \alpha_{\mu j} S_j^*}{R_\mu^*} = \dfrac{l_0 \left(1 + \epsilon_l \tilde{l}_\mu\right) + \alpha_0 S_0 \left(N_S + \epsilon_\alpha \sum_j \tilde{\alpha}_{\mu j} + \epsilon_\alpha \epsilon_S \sum_j \tilde{\alpha}_{\mu j} \tilde{s}_j\right)}{R_0 \left(1 + \epsilon_R \tilde{r}_\mu\right)} & (2.105) \\ -\gamma_{j\mu} R_\mu^* + \alpha_{\mu j} = -\gamma_0 R_0 \left(1 + \epsilon_\gamma \tilde{\gamma}_{j\mu} + \epsilon_R \tilde{r}_\mu + \epsilon_\gamma \epsilon_R \tilde{\gamma}_{j\mu} \tilde{r}_\mu\right) + \alpha_0 \left(1 + \epsilon_\alpha \tilde{\alpha}_{\mu j}\right) & (2.106) \\ \sigma_{i\nu} \gamma_{i\nu} S_i^* = \sigma_0 \gamma_0 S_0 \left(1 + \epsilon_\sigma \tilde{\sigma}_{i\nu} + \epsilon_\gamma \tilde{\gamma}_{i\nu} + \epsilon_S \tilde{s}_i + \epsilon_\sigma \epsilon_\gamma \tilde{\sigma}_{i\nu} \tilde{\gamma}_{i\nu} + \epsilon_\sigma \epsilon_S \tilde{\sigma}_{i\nu} \tilde{s}_i \right. & \\ \quad \left. + \epsilon_\gamma \epsilon_S \tilde{\gamma}_{i\nu} \tilde{s}_i + \epsilon_\sigma \epsilon_\gamma \epsilon_S \tilde{\sigma}_{i\nu} \tilde{\gamma}_{i\nu} \tilde{s}_i\right) & (2.107) \end{cases}$$

Standard deviation expansion at first order

We assume the relative standard deviations are small:

$$\epsilon_P \ll 1, \ \forall P \in \{l_\nu, R_\nu^*, S_i^*, \gamma_{i\nu}, \alpha_{\nu i}, \sigma_{i\nu}\}. \tag{2.108}$$

The previous equations can be rewritten as:

$$\begin{cases} \dfrac{l_\mu + \sum_j \alpha_{\mu j} S_j^*}{R_\mu^*} = \dfrac{l_0 + N_S \alpha_0 S_0}{R_0} - \epsilon_R \dfrac{l_0 + N_S \alpha_0 S_0}{R_0} \tilde{r}_\mu + \epsilon_l \dfrac{l_0}{R_0} \tilde{l}_\mu + \epsilon_\alpha \dfrac{\alpha_0 S_0}{R_0} \sum_j \tilde{\alpha}_{\mu j} & (2.109) \\ -\gamma_{j\mu} R_\mu^* + \alpha_{\mu j} = -\gamma_0 R_0 + \alpha_0 - \epsilon_\gamma \gamma_0 R_0 \tilde{\gamma}_{j\mu} - \epsilon_R \gamma_0 R_0 \tilde{r}_\mu + \epsilon_\alpha \alpha_0 \tilde{\alpha}_{\mu j} & (2.110) \\ \sigma_{i\nu} \gamma_{i\nu} S_i^* = \sigma_0 \gamma_0 S_0 \left(1 + \epsilon_\sigma \tilde{\sigma}_{i\nu} + \epsilon_\gamma \tilde{\gamma}_{i\nu} + \epsilon_S \tilde{s}_i\right) & (2.111) \end{cases}$$

where we neglect all terms of order $\mathcal{O}(\epsilon^2)$. We now assume furthermore that the relative standard deviations of every parameter in the model more or less have the same value which again is assumed small, *i.e.* we set :

$$\epsilon_P \approx \epsilon \ll 1, \ \forall P \in \{l_\nu, R_\nu^*, S_i^*, \gamma_{i\nu}, \alpha_{\nu i}, \sigma_{i\nu}\}. \tag{2.112}$$

This allows us to rewrite the jacobian at equilibrium as:

$$J^* = J_0 + \epsilon \tilde{J}, \tag{2.113}$$

with

$$J_0 = \begin{pmatrix} -\Delta_0 & \Gamma_0 \\ B_0 & 0 \end{pmatrix} \tag{2.114}$$

where

$$(\Delta_0)_{\mu\nu} \equiv \Delta_0 \delta_{\mu\nu} = \frac{l_0 + N_S \alpha_0 S_0}{R_0} \delta_{\mu\nu} \tag{2.115}$$

$$(\Gamma_0)_{\mu i} \equiv \Gamma_0 = -\gamma_0 R_0 + \alpha_0 \tag{2.116}$$

$$(\mathsf{B}_0)_{i\nu} \equiv \mathsf{B}_0 = \sigma_0 \gamma_0 S_0 \tag{2.117}$$

and

$$\tilde{J} = \begin{pmatrix} \tilde{\Delta} & \tilde{\Gamma} \\ \tilde{\mathsf{B}} & 0 \end{pmatrix} \tag{2.118}$$

with

$$\tilde{\Delta}_{\mu\nu} \equiv \left(-\Delta_0 \tilde{r}_\mu + \frac{l_0}{R_0} \tilde{l}_\mu + \frac{\alpha_0 S_0}{R_0} \sum_j \tilde{\alpha}_{\mu j} \right) \delta_{\mu\nu} \tag{2.119}$$

$$\tilde{\Gamma}_{\mu i} \equiv \alpha_0 \tilde{\alpha}_{\mu i} - \gamma_0 R_0 \left(\tilde{\gamma}_{i\mu} + \tilde{r}_\mu \right) \tag{2.120}$$

$$\tilde{B}_{i\mu} \equiv B_0 \left(\tilde{\sigma}_{i\mu} + \tilde{\gamma}_{i\mu} + \tilde{s}_i \right) \tag{2.121}$$

Using Jacobi's formula [34], the equation $\det\left(J^* - \lambda \mathbb{1}_{N_R+N_S}\right) = 0$ can be rewritten as:

$$\boxed{\det(J_0 - \lambda \mathbb{1}_{N_R+N_S} + \epsilon J^*) = \det(J_0 - \lambda\ \mathbb{1}_{N_R+N_S}) + \epsilon\ \mathrm{Tr}\left(\mathrm{adj}(J_0 - \lambda\ \mathbb{1}_{N_R+N_S})\tilde{J}\right) = 0.} \tag{2.122}$$

where adj $(\dots)$ is the adjugate operator (*i.e.* which yields the transpose of the cofactor matrix). This equation is a complicated polynomial of degree $N_R + N_S$. We do not know how to find an easily computable solution for $\epsilon > 0$. An explicit solution can however be computed when $\epsilon = 0$.

Zero variance case When $\epsilon = 0$, Eq.(2.122) becomes:

$$\det(J_0 - \lambda \mathbb{1}_{N_R+N_S}) = \det \begin{pmatrix} -\Delta_0 - \lambda \mathbb{1}_{N_R} & \Gamma_0 \\ B_0 & -\lambda \mathbb{1}_{N_S} \end{pmatrix} = 0 \tag{2.123}$$

If we assume that $\lambda \neq 0$, using a reasoning similar to Section 2.2.2 we can write the previous equation as:

$$\det\left(\lambda^2 \mathbb{1}_{N_R} + \Delta_0 \lambda - \Gamma_0 \mathsf{B}_0\right) = 0 \tag{2.124}$$

Component-wise, we have:

$$\left(\lambda^2 \mathbb{1}_{N_R} + \Delta_0 \lambda - \Gamma_0 \mathsf{B}_0\right)_{\mu\nu} = \left(\lambda^2 + \Delta_0 \lambda\right) \delta_{\mu\nu} - \Gamma_0 \mathsf{B}_0. \tag{2.125}$$

Appendix 5.1.1 explains how to find the non-zero solutions of Eq.(2.124). They are given by the roots of the polynomial:

$$(\lambda + \Delta_0)^{N_R - 1} \left(\lambda^2 + \Delta_0 \lambda - N_R \Gamma_0 \mathsf{B}_0\right) = 0. \tag{2.126}$$

That equation gives us $N_R - 1 + 2 = N_R + 1$ non-zero eigenvalues, which means that there are $N_S - 1$ zero eigenvalues. The two eigenvalues different from $-\Delta_0$ or 0 are the roots of the second degree polynomial:

$$\lambda^2 + \Delta_0 \lambda - N_R \Gamma_0 \mathrm{B}_0 = 0. \tag{2.127}$$

In the end, the spectrum is given by:

- if $\Gamma_0 < -\frac{\Delta_0^2}{4N_R \mathrm{B}_0}$:

$$\sigma(J_0) = \left\{ 0, \dots, 0, -\Delta_0, \dots, -\Delta_0, -\frac{\Delta_0}{2}\left(1 \pm i\sqrt{-\left(1 + \frac{4N_R\Gamma_0\mathrm{B}_0}{\Delta_0^2}\right)}\right) \right\} \tag{2.128}$$

- if $\Gamma_0 = -\frac{\Delta_0^2}{4N_R \mathrm{B}_0}$:

$$\sigma(J_0) = \left\{ 0, \dots, 0, -\Delta_0, \dots, -\Delta_0, -\frac{\Delta_0}{2}, -\frac{\Delta_0}{2} \right\} \tag{2.129}$$

- if $\Gamma_0 > -\frac{\Delta_0^2}{4N_R \mathrm{B}_0}$:

$$\sigma(J_0) = \left\{ 0, \dots, 0, -\Delta_0, \dots, -\Delta_0, -\frac{\Delta_0}{2}\left(1 \pm \sqrt{1 + \frac{4N_R\Gamma_0\mathrm{B}_0}{\Delta_0^2}}\right) \right\} \tag{2.130}$$

It then becomes clear that the system is dynamically unstable if and only if $\frac{4N_R\Gamma_0\mathrm{B}_0}{\Delta_0^2} > 0$. Because $N_R, \mathrm{B}_0 > 0$, we get the condition:

$$\boxed{\text{The non-variance system is dynamically unstable } \iff \Gamma_0 > 0.} \tag{2.131}$$

If $\Gamma_0 \leqslant 0$, the fully connected system will be marginally stable. Note that the RHS of the feasibility condition Eq.(2.28) is equivalent in the fully connected case to :

$$\alpha_0 \lessapprox \min(1-\sigma_0, \sigma_0)\gamma_0 R_0 \iff \Gamma_0 \lessapprox \left[\min(1-\sigma_0, \sigma_0) - 1\right]\gamma_0 R_0 < 0, \tag{2.132}$$

which means that the case $\Gamma_0 > 0$ is simply not feasible. So in the end the feasible fully connected case is marginally stable and its non-zero eigenvalues have a negative real part.

We see that all but one non-zero eigenvalues are given by $-\Delta_0$. However the last eigenvalue is determined by $4N_R\Gamma_0\mathrm{B}_0 + \Delta_0^2$. It gives us the "deviation" away from $-\Delta_0/2$ and hence also plays an essential role in the stability. Looking at its sign allows us to find possibly more locally dynamically stable zones of the metaparameters space $\mathcal{M}$. Indeed we expect to find more stable systems in the case of Eq.(2.130) if $4N_R\Gamma_0\mathrm{B}_0 + \Delta_0^2$ is small, *i.e.*

$$4N_R\Gamma_0\mathrm{B}_0 + \Delta_0^2 \ll 1 \iff 4N_R\sigma_0\gamma_0 S_0(\alpha_0 - \gamma_0 R_0) + \frac{l_0^2}{R_0^2} + \frac{2N_S\alpha_0 S_0 l_0}{R_0^2} + \frac{N_S^2\alpha_0^2 S_0^2}{R_0^2} \ll 1. \tag{2.133}$$

This equation will be useful in order to predict which metaparameters will lead to dynamically stable systems (see Section 3.2.4).

2.3 Structural stability

When studying dynamical stability, we investigate what happens when the equilibria abundances $\{R_\mu^*, S_i^*\}$ of a given equilibrium point are perturbed. The question of *structural stability* looks also at the behaviour of a given system when perturbed away from equilibrium. However, structural stability focuses on the perturbations of the parameters of the model *i.e.* $\{l_\mu, m_\mu, \gamma_{i\mu}, \alpha_{\mu i}, \sigma_{i\mu}, d_i\}$. Namely we will try to answer the following question :

Given an equilibrium point, does the system go back to a positive-valued equilibrium when some of the model parameters are changed? If yes, how much can they be changed before the system evolves in such a way that it does not reach a positive-valued equilibrium?

2.3.1 Definitions

Studying how a system responds to the perturbation of all parameters $\{l_\mu, m_\mu, \gamma_{i\mu}, \alpha_{\mu i}, \sigma_{i\mu}, d_i\}$ is a quite difficult problem. So we will try to simplify it by perturbing *only one* parameter. We make the somewhat arbitrary choice of perturbing the external feeding rate l_μ, since it is essentially the only parameter one can control experimentally [**is this true?**]. More precisely, consider $\Delta_S \in [0,1]$. We say that a given system $p \in \mathcal{P}$ is *structurally stable* under the perturbation Δ_S, if under the transformation

$$l_\mu \to \hat{l}_\mu \equiv l_\mu \left(1 + \Delta_S \nu_\mu\right) \tag{2.134}$$

the transformed set of parameters $\left\{\hat{l}_\mu, m_\mu, \gamma_{i\mu}, \alpha_{\mu i}, \sigma_{i\mu}, d_i\right\}$ gives rise under time evolution to a positive valued-equilibrium $\left\{\hat{R}_\mu^*, \hat{S}_i^*\right\}$. In the equation above, ν_μ is a random variable drawn from a uniform distribution of support $[-1,1]$. In words, we start with an initial parameters set at an equilibrium point, which is constant under time evolution, and see how much we can change the resources external feeding rate until some consumers start to die out as the new system is time-evolved.

Similarly to what was done for feasibility and dynamical stability, we can define the *parameters set structural stability function* $\mathfrak{S} : [0,1] \times \mathcal{P} \to \{0,1\}$ in the following way $\forall \Delta_S \in [0,1], p \in \mathcal{P}$:

$$\mathfrak{S}(\Delta_S, p) = \begin{cases} 1 \text{ if } p \text{ is structurally stable under the perturbation } \Delta_S, \\ 0 \text{ otherwise.} \end{cases} \tag{2.135}$$

For a fixed p, we expect $\mathfrak{S}(\Delta_S, p)$ to behave as a step function of Δ_S : we may only perturb the parameters so much before they suddenly become structurally unstable.

The corresponding metaparameters set function, the *metaparameters set structural stability function* $\mathcal{S}$ can also be defined as the function which, given a set of metaparameters and a consumption-syntrophy couple of binary matrices, tells you how probable it is that you draw a system structurally stable under a perturbation Δ_S. Mathematically, $\mathcal{S} : [0,1] \times \mathcal{M} \times \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S} \to [0,1]$ is defined $\forall \Delta_S \in [0,1], m \in \mathcal{M}, B = (G, A) \in \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S}$:

$$\boxed{\mathcal{S}(\Delta_S, m, B) = \text{Probability}\left\{\mathfrak{S}(\Delta_S, \mathcal{A}(m, B)) = 1\right\}} \tag{2.136}$$

Because we expect a step-like drop of $\mathfrak{S}$ as Δ_S increases, we expect also a somewhat sharp drop from $\mathcal{S} \approx 1$ to $\mathcal{S} \approx 0$. To quantify this, one can define the *critical structural perturbation* $\Delta_S^*(m, G, A)$ of a consumption-syntrophy network couple implicitly as :

$$\mathcal{S}(\Delta_S^*(m, G, A), m, G, A) = 0.5 \tag{2.137}$$

Methods 2.3.2 below explains how $\Delta_S^*(m, G, A)$ can be estimated numerically.

2.3.2 Numerical estimate of the critical structural perturbation

As explained above, the critical structural perturbation of set of metaparameters m and a consumption-syntrophy couple of matrices is the point where we shift from structural stability to instability. In that sense, $\Delta_S^*(m, G, A)$ is a measure of how good (m, G, A) respond to structural perturbation and can be interpreted geometrically as the radius of a sphere of "tolerance" around m **Shouldn't we change all metaparameters in order to achieve this?**. It turns out that Δ_S^* can be estimated numerically quite easily.

Indeed, to decide whether a system is structurally stable or not, one can simply perturb the parameters of the system, let it time-evolve until it reaches a new equilibrium and count how many of the original consumers are still present at the new equilibrium. By repeating this procedure many times one gets a good estimate of the *probability of observing an extinction* $P_E(\Delta_S, m, G, A)$ after a structural perturbation Δ_S and it is clear that

$$\mathcal{S}(\Delta_S, m, G, A) = 1 - P_E(\Delta_S, m, G, A). \tag{2.138}$$

So $\Delta_S^*(m, G, A)$ can be found by computing $P_E(\Delta_S, m, G, A)$ over the range $[0, 1]$ and finding at which point it is equal to 0.5. In practice, a general solver involving methods from the C++ `GSL` library was implemented in order to get a good estimate on $\Delta_S^*(m, G, A)$ (see Appendix 5.2.7).

3 Results

After establishing in the previous chapter the framework and methods we will use, we now work on getting results that are both biologically relevant and interpretable. We keep the same structure as Chapter 2 and will focus first on the crucial but sometimes overlooked[18] question of *feasibility*, which aims to determine what regimes of parameters and meta-parameters give rise to microbial communities that could exist in Nature. We will then move on to study how microbial communities resist to perturbations, *i.e.* how stable they are. Two types of perturbations will be considered. Determining how communities react to perturbations of the resources and microbial species abundances will lead us to consider their *local dynamical stability*. Finally, we will study how they react to environmental perturbations, *i.e.* we will quantify their *structural stability*.

TO DO : put the following in the future "global aims" section The aim of this thesis is to study how these quantities change with respect not only to the consumption matrix γ, but also to the syntrophy matrix α. As explained above (Introduction 1.2.2) the large complexity of this task made us move to a statistical approach, where general matrices are separated in two parts, namely average non-zero link strength (what we called metaparameter) and network structure. That led to the definition of the binary syntrophy network matrix A. We here will need to simplify our approach even more.

Indeed our goal is to study stability properties of many consumption-syntrophy networks (G, A). We decided to focus on a set of G-matrices varying two characteristics of G, namely its connectance κ_G and nestedness (which, for the case of G, we call ecological overlap) η_G. By symmetry we should also choose a set of A's with different connectances and nestedness and study every possible pair (G, A) together, making the total number of microbial communities to study equal to the multiplication of the number of consumption networks by the number of syntrophy networks. Working that way, although more scientifically thorough, would demand more time than what is allowed for this type of Thesis and is hence not possible. Because we still would like to study the effect of the shape of A, we decide, for each consumption matrix G, to consider four A "scenarios"[19]:

- "Fully Connected" (FC): A is filled with ones only, $A_{\mu i} = 1$. This corresponds to a so-called "mean-field" approximation. Every consumer releases every resource at the same intensity (up to some noise) .

- "No Intraspecific Syntrophy" (NIS): the structure of A is such that consumers are not allowed to release what they consume, *i.e.* there is no coprophagy. Apart from this, they release everything else. G-matrices with a small connectance will have an A-matrix with a large connectance (not far away from the FC case) and vice-versa.

[18] The "traditional" approach over the course of years (see *e.g.* [24]) to the study of ecological communities has been to study local dynamical stability through the means of random matrix theory. The idea is to study how the statistical properties of the jacobian matrix at equilibrium J^* influence stability, with the assumption that J^* is a random matrix. Only recently**Is this true?** have discussions arised [35, 36] about in what limit, if any, that assumption is a good representation of Nature, especially since we discovered ecological communities are shaped by dynamics which may lead to non particular, non-random, structures [22, 37, 38, 39].

[19] We talk here about scenarios because, apart from the FC case, A will generally be different for each G. It is important to remember that even though we compare A-matrices "in the same scenario", those may have a very different shape.

- "Low Resource Interaction" (LRI): A is the outcome of the MCMC algorithm[20] described in Methods 2.2.4. The purpose of this algorithm is to build an A that minimizes the energy $E(G, A)$ (Eq.2.93), and hence get the A which for a given G, brings the system as close to satisfaction of Eq.(2.76) as possible. The connectance of the A-matrix is taken as the one of the G-matrix.
- "Random Structure" (RS) : A is taken as a random matrix (with the right dimensions), which has the same connectance as the corresponding G but where non-empty links have been randomly placed.

These four scenarios represent very different regimes, which, we hope, will have an impact on the feasibility or stability of the microbial communities considered.

3.1 Feasibility

As explained in Methods 2.1, before addressing the question of the stability of a system, be it dynamical or structural, it is important to study whether that system is *feasible*. In short we must answer the question: "does it make sense to talk about this system? What are the restrictions on the parameters of a microbial community in order for it to exist?". Methods 2.1.1 provides a mathematical definition of *feasibility*. Intuitively, we say that in order to be feasible, a microbial community should respect two conditions: biomass must be conserved and its parameters must have a direct biological interpretation. Methods 2.1.2 defines $\forall x \in [0, 1]$ the x-feasibility region $\mathcal{F}_x^{G,A} \subset \mathcal{M}$ of the consumption-syntrophy network $(G, A) \in \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S}$. The idea behind that concept is that every metaparameter set $m \in \mathcal{M}$ contained in the x-feasibility region $\mathcal{F}_x^{G,A}$ gives rise to a percentage x of feasible systems. Knowing where the different $x-$feasibility regions are located allows us to predict relationships between the different parameters of microbial communities. In particular, we would like to know which metaparameters $(\gamma_0, S_0, \alpha_0)$ yield through the algorithmic procedure $\mathcal{A}$ a hundred percent of the time a feasible parameters set. In short, we are interested in $\mathcal{F}_1^{G,A}$. A direct study of every triplet $(\gamma_0, S_0, \alpha_0)$ is difficult so we take a more "implicit" path in which we determine, as a function of α_0, what values (γ_0, S_0) can take such that $m = (\gamma_0, S_0, \alpha_0)$ is in $\mathcal{F}_1^{G,A}$.

3.1.1 The feasibility region in the absence of syntrophy

We start with the "null" case and study feasibility in the absence of syntrophy. A first order approximation of the fully feasible volume $\mathcal{F}_1^{G,A}$ is given by Eq.(2.28). In the absence of syntrophy $\alpha_0 = 0$, it becomes:

$$\gamma_0 R_0 \lessapprox \frac{l_0}{\max_\nu\{\deg(G, \nu)\} S_0}. \tag{3.1}$$

That relation provides a lot of biological insight about how the different metaparameters of actual microbial communities should be linked together:

[20]Note that we will take a constant value α_0 (given in Methods 2.2.4) and $\gamma_0 = 0.2$. A more thorough analysis should build the optimal LRI matrix *corresponding to each* (γ_0, α_0). That would take too much time which is why we decided to keep γ_0 and α_0 constant.

- For a fixed consumption interaction (*i.e.* constant γ_0 and G) and average consumers abundance S_0, increasing the average resource equilibrium abundance R_0 implies increasing the external resource input rate. This is completely expected: since the microbes consume some part of the resources, more available resources can only come from outside in the absence of syntrophy.

- For a fixed consumption network G, resources abundance R_0 and resource input rate l_0, a larger range of possible consumption rate γ_0 is possible if the consumers abundance is decreased. Since there are fewer the consumers, they can individually eat more from the available resources.

- For every metaparameter except γ_0 fixed, arranging the shape of G such that few consumers eat from the same resources[21] increases the range of possible consumption rates. If the total amount of resources is fixed, consumers can individually eat more of it if they do not have to share it!

Equation (3.1) can be confronted to simulations. The first step is to note that for all the matrices in the set we chose **TO DO: give a name to the matrix set**, there exists a fully feasible zone, *i.e.* $\mathcal{F}_1^{G,0}(\alpha_0 = 0) \neq \{\}\ \forall G$. There exists an overlap between all these regions $\mathcal{F}_1^{S_{25}}(0) \equiv \bigcap_{G\in G_{25}} \mathcal{F}_1^{G,0} \neq \{\}$, such that the critical feasibility $f^*(S) = 1$ and the critical feasibility region at no syntrophy is the common fully feasible volume $\mathcal{F}_1^{G_{25}}$ (see Eqs.(2.11) and (2.12)) **TO DO: add figure of this?**.

Figure 3.1.1 shows the typical proportion of feasible systems without syntrophy for two consumption matrices G_1 and G_2 in the set G_{25}. We generally observe two distinct zones in the (γ_0, S_0) space: full feasibility ($\mathcal{F}((\gamma_0, S_0, 0), G, 0)$ as defined in Eq.(2.5) is equal to 1) and full infeasibility ($\mathcal{F}((\gamma_0, S_0, 0), G, 0) = 0$). These two zones are separated by a narrow region of partial feasibility $0 < \mathcal{F}((\gamma_0, S_0, 0), G, 0) < 1$. Since that region is very thin, we can define the "boundary" line between feasibility and infeasibility as the level[22] $\mathcal{F}((\gamma_0, S_0, 0), G, 0) = 0.5$. Theoretically, that sharp transition between feasibility and infeasibility happens when both sides of the inequality (3.1) are equal, *i.e.* at $\gamma_0 R_0 = l_0/\max_\nu\{\deg(G,\nu)\}S_0$. Hence we expect the boundary measured above to be well characterized by the curve $S_0 = K\gamma_0^{-1}$ with $K = l_0/(R_0 \max_\nu \{\deg(G,\nu)\})$.

For G_1, the theoretical expectation is $S_0 = 0.125\gamma_0^{-1}$ and a fit on the numerical results gives $S_0 = (0.124 \pm 3 \times 10^{-8})\gamma_0^{-1}$ so the theoretical estimate is very accurate. For G_2, we expect $S_0 = 0.077\gamma_0^{-1}$. A fit gives $S_0 = (0.076 \pm 7 \times 10^{-9})\gamma_0^{-1}$. Again, the theoretical value is very close to the numerical measurement. However, the numerical estimate does not always match the theoretical value that well. Figure 3.1.2 shows the relative error $\Delta_K \equiv 1 - K_{\text{theorical}}/K_{\text{fit}}$. We see that in general $\Delta_K < 0$, *i.e.* the theoretical expectation tends to overestimate the fully feasible region. This is probably due to noise (*i.e.* coming from the difference between the metaparameters and the parameters) in the actual systems and the topology of the consumption matrix G. Figure 3.1.2 shows that the lower the ecological overlap or connectance of G, the worse the theoretical estimate. Finding a more accurate

[21] Indeed, $\deg(G,\nu)$ is the number of species that consume resource ν. Reducing the number of consumers that eat from the same species reduces $\max \deg(G,\nu)$ in Equation (3.1).

[22] Numerically because of the possible noise, we take as part of the boundary every (γ_0, S_0) such that $0.4 < \mathcal{F}((\gamma_0, S_0, 0), G, 0) < 0.6$.

approximation which takes into account the deviations away from the metaparameters and the topology of G remains a challenge for future studies.

We can similarly measure the common fully feasible volume in the absence of syntrophy $\mathcal{F}_1^{G_{25}}(0)$ (depicted in Fig.3.1.3), which according to Eq.(3.1) is inversely proportional to the largest maximal row degree of the matrix set. For the set S_{25}, this yields in theory: $S_0 = 0.053\gamma_0^{-1}$. A fit on the points at the edge yields the critical boundary $S_0 = (0.042 \pm 10^{-8})\gamma_0^{-1}$, which is $\sim 21\%$ away from the theoretical prediction. The discrepancy is probably due to the difference between the way we estimate the boundary numerically and analytically.

3.1.2 Impact of syntrophy on the feasible region

Above we computed feasible volumes when there is no syntrophy *i.e.* $\mathcal{F}_1^{G,0}$. Because there was no syntrophy, we did not need to specify what the structure of A was. The next naturally arising question is then: what happens to the feasible region of a community with a given consumption matrix G when we add a syntrophic interaction, *i.e.* a syntrophic network A of average interaction strength α_0? More precisely, how does the shape of $\mathcal{F}_1^{G,A}$ change as a function of A and α_0?

A layer of complexity arises on top of the problem discussed above: apart from the structure of the consumption matrix G, we now have to think about both the topology of A and also α_0 as well. As explained at the beginning of this section, studying in detail the topology of A is too large of a scope for this study, so we focus on the four A cases enunciated above (FC, NIS_1, LRI_1, RS_1). Concerning the question of α_0, we would like to be "fair" and study syntrophy strengths that are feasible for all the consumption-syntrophy networks considered $(G, A) \in S_{25}$. The largest common fully feasible syntrophy $\alpha_C^F(S_{25})$, which is the value of α_0 such that $\mathcal{F}_1^{G,A} \neq \{\}\ \forall (G, A) \in S_{25}$, can be estimated with the help of Eq.(2.28):

$$\alpha_0 \lessapprox \frac{\min(1-\sigma_0, \sigma_0)\gamma_0 R_0}{\max_{(G,A)\in S}\left\{\max_i\left\{\frac{\deg(A,i)}{\deg(G,i)}\right\}\right\}} \approx 0.01\gamma_0 \leqslant 0.01. \tag{3.2}$$

We will hence investigate the impact of α_0 evaluated at ten different values from 0 to 0.015. Since we do not change α_0 in a continuous way but instead focus on different "α_0-slices" of $\mathcal{F}_1^{G,A}$, the object of our attention is the set of (γ_0, S_0) such that $(\gamma_0, S_0, \alpha_0) \in \mathcal{F}_1^{G,A}$. In a rather unfortunate notation, we will refer to it as $\mathcal{F}_1^{G,A}(\alpha_0)$:

$$\mathcal{F}_1^{G,A}(\alpha_0) \equiv \left\{(\gamma_0, S_0) : (\gamma_0, S_0, \alpha_0) \in \mathcal{F}_1^{G,A}\right\}. \tag{3.3}$$

Because Equation (2.28) depends on the structure of G and of A, we expect $\mathcal{F}_1^{G,A}(\alpha_0)$ to depend heavily on the topology of both the consumption and syntrophy matrices.

The influence of matrix topology

Figure 3.1.4 shows that indeed $\mathcal{F}_1^{G,A}(\alpha_0)$ changes significantly not only with syntrophy α_0 but also with the network structure of the consumption matrix G. How $\mathcal{F}_1^{G,A}(\alpha_0)$ changes as a function of A is a more difficult question. **TO DO: explain that this will be studied later : put reference to new figure (how do df and alpha0 change from the null case)?**

We observe a general trend among all matrices: as syntrophy increases, the fully feasible region moves horizontally to the right towards a higher γ_0. This can be explained with Eq.(2.28) which provides a lower bound to γ_0 when $\alpha_0 > 0$. Note that taking into account $\alpha_0 > 0$ does not bound S_0, such that at a fixed γ_0, every S_0 from 0 to the upper boundary critical curve $\sim \gamma_0^{-1}$ discussed before remains a fully feasible point. So in general, as syntrophy increases, systems with a high consumption rate and a low consumers abundance at equilibrium will remain feasible, while other simply will not exist. This makes sense intuitively:**TO DO: put explanation here**

Because of that shift to the right, $\mathcal{F}_1^{G,A}(\alpha_0)$ shrinks in size[23]: as syntrophy is increased, the set of possible consumption rate and average consumers abundance is more restricted. Figure 3.1.5 shows that typically the feasibility volume Vol $\left(\mathcal{F}_1^{G,A}(\alpha_0)\right)$, formally defined in Appendix 5.2.6, decays in an exponential-like fashion as α_0 increases. This is a simple translation of the biological hindsight that, because of the physical considerations discussed above, the number of systems that can sustain a given syntrophic interaction strength decreases with that interaction strength, *i.e.* no system can support an arbitrarily large syntrophy. The shrinkage of the feasibility volume can be quantified by defining the *feasibility decay rate* d_F, which is obtained by performing a non-linear regression to find the coefficients $c_1, c_2, d_F \in \mathbb{R}^+$ that satisfy best[24]:

$$\text{Vol}\left(\mathcal{F}_1^{G,A}(\alpha_0)\right) \approx c_1 \exp\left(-d_F \alpha_0\right) - c_2. \tag{3.4}$$

The value of $d_F(G, A)$ tells us how fast the feasible volume shrinks for a given consumption-syntrophy couple (G, A). In that sense $d_F(G, A)$ provides a measure of how good (G, A) can sustain an increase in syntrophy[25]. If d_F is low then the system can bear an increase of syntrophy and stay feasible. On the opposite, if d_F is high, if syntrophy is increased too much, the microbial community will have to fundamentally change (*e.g.* move to a (G, A) configuration with a lower d_F) or it will die out.

Figure 3.1.7 shows how $d_F(G, A)$ changes as a function of the structure of G, for different types of A. Interestingly, the structure of A does not provide any real difference, except for G with $\eta_G = 0.15$ and $\kappa_G = 0.18$. For that specific matrix, the LRI scenario for A reduced by a factor of three the decay rate, compared to the fully connected case. That result is only true for this specific matrix and does not hold for the others, where $d_F(G, A)$ seems to almost not depend on A. Furthermore a very strong trend can be observed, for all consumption matrices and all structures of A: for a given connectance κ_G, d_F is increased if the ecological overlap η_G is increased and for a given ecological overlap, d_F decreases if the connectance is increased.**TO DO: correct this** In biological terms, systems where there

[23]Note that when we talk about the size of $\mathcal{F}_1^{G,A}(\alpha_0)$, it is understood relatively to the $(\gamma_0, S_0) \in [0, 1]^2$ unit square.

[24]In practice, standard functions of the `numpy` and `scipy` Python libraries are used to perform that fitting procedure.

[25]One could also desire to define a *critical feasible syntrophy* as the smallest syntrophy that gives a zero fully feasible volume. This would be a very interesting quantity to study and could be easily found as the root of the RHS of Eq.(3.4). We tried doing this, because the errors on d_F and on the two other fitting coefficients are already quite large (see the caption of Fig.3.1.7), the errors on the critical feasible syntrophy we obtained were way too large, making our results essentially meaningless.

is a small ecological overlap but a lot of links in the food consumption network will be favoured. Microbial communities where consumers on average eat from a lot of different resources (*i.e.* each their own) can sustain a larger syntrophic interaction than others.

Common fully feasible volume

After studying the effect of syntrophy on each consumption-syntrophy network (G, A), we can focus on the overlap of all the fully feasible regions: the common fully feasible region $\mathcal{F}_1^{S_{25}} \equiv \bigcap_{(G,A)\in S_{25}} \mathcal{F}_1^{G,A}$. Similarly to above, we slightly abuse notation and write:

$$\mathcal{F}_1^{S_{25}}(\alpha_0) \equiv \left\{(\gamma_0, S_0) : (\gamma_0, S_0, \alpha_0) \in \mathcal{F}_1^{S_{25}}\right\}. \tag{3.5}$$

Figure 3.1.8 shows the evolution of $\mathcal{F}_1^{S_{25}}(\alpha_0)$ as syntrophy increases. In accordance to what was observed individually for each matrix before, the RS scenario allows for a larger syntrophy: at a given α_0, there are more feasible (γ_0, S_0) points if A has a random structure. The three other scenarios do not offer a significant difference and at $\alpha_0 = 9.1 \times 10^{-3}$ there are no more (γ_0, S_0) that are fully feasible for all matrices. This means that the largest common fully feasible syntrophy respects $7.8 \times 10^{-3} < \alpha_C^F(S_{25}) \leqslant 9.1 \times 10^{-3}$ for those three scenarios, which puts the estimate $\alpha_C^F(S_{25}) \approx 0.01$ of Eq.(3.2) not far from reality.

Without surprise, Vol $\left(\mathcal{F}_1^{S_{25}}(\alpha_0)\right)$ also decays in an exponential-like fashion (Figure 3.1.9). An exponential fit allows to find the feasibility decay rates of each scenario: for the FC case, $d_F = 302 \pm 18$ (in units of syntrophy α_0); for NIS, $d_F = 297 \pm 17$; for LRI, $d_F = 291 \pm 17$ and finally for RS, $d_F = 115 \pm 30$. The FC, NIS and LRI scenarios of A do not produce a significant difference in feasibility. However a random A-matrix (RS scenario) allows for a 2.5 times smaller feasibility decay rate and hence at a given syntrophy α_0, for much more pairs of common feasible (γ_0, S_0). Without any surprise as well we observe the same shift of $\mathcal{F}_1^{S_{25}}$ towards points with a high γ_0 and consequently a small S_0. Overall, independently of the consumption-syntrophy network structure, lowly abundant microbial communities which eat aggressively can sustain a larger syntrophy than others. Among these, the ones which have a random syntrophy network are the most "compatible" with even larger syntrophic interaction.

The influence of the matrix dimensions

We focused above on microbial communities with the same number of consumers and resources: $N_R = N_S = 25$. But such systems lie at what has been referred to in the literature as May's stability bound [29], which is defined as an ecological community where the number of resources is equal to the number of species. According to the competitive exclusion principle[26], an ecological system which has as many resources as consumers is the border case where coexistence, *i.e.* feasibility in the terms used here, starts to exist. In a way, the study conducted before can be seen as a borderline case and it can be very fruitful to investigate the behaviour of systems where the number of resources has been increased to $N_R = 50$.

[26]The heavily debated and often misunderstood [40] *competitive exclusion principle*, also known as Gause's principle, states that "Complete competitors cannot coexist" [40], or more generally that "the number of consumer species in steady coexistence cannot exceed that of resources" [41].

Figure 3.1.10 is the $N_R = 50$ equivalent to Figure 3.1.4. We observe that increasing N_R has a non-trivial effect on feasibility, and that effect differs for each consumption-syntrophy network. For instance, for G with $\eta_G = 0.6$ and $\kappa_G = 0.32$, adding more resources increases the maximal syntrophy bearable by the system (Fig.3.1.10a compared to Fig.3.1.4a). On the contrary, for G with $\eta_G = 0.15$ and $\kappa_G = 0.12$, it decreases it from 1.4×10^{-2} to 7.8×10^{-3}.

Figure 3.1.11 shows the common feasibility region $F_1^{S_{50}}$ of the $N_R = 50$ matrix set. We see that, compared with the $N_R = 25$ case (Fig.3.1.8), an overall lesser syntrophy can be achieved: the largest common fully feasible syntrophy is reduced **TO DO: put prediction here**. Finally Figure 3.1.12 shows that indeed it is really hard to predict the effect that increasing the number of resources will have on a specific consumption-syntrophy network. Indeed d_F, which we use to measure how big of a syntrophy a consumption-syntrophy network (G, A) can bear, does not have a clear pattern, at least not under the matrix metrics we chose to measure. It is a sign that this question needs a further and deeper investigation. **TO DO: rework last paragraph and more importantly put figure of dF(NR=50)/dF(NR=25) as a function of eta and kappa**

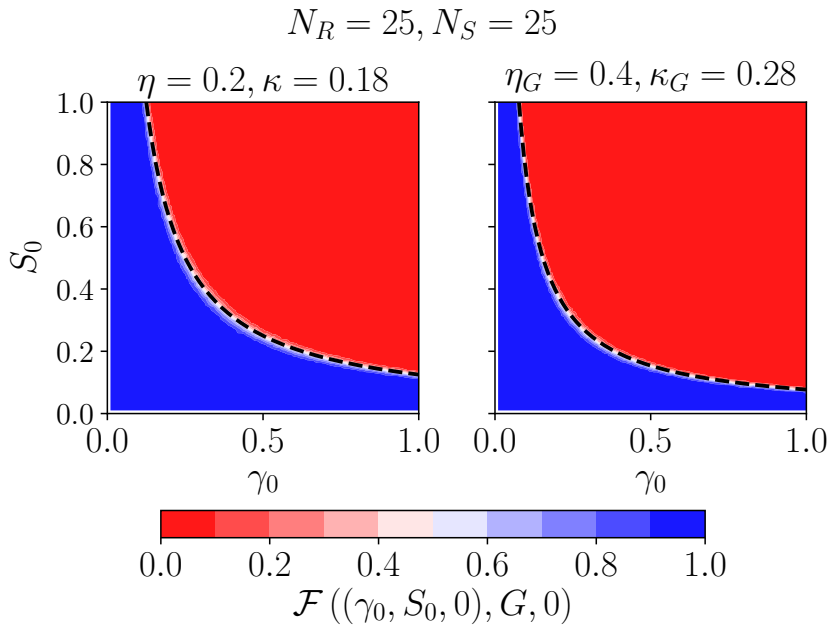

Figure 3.1.1: Plot of the feasibility region in the absence of syntrophy. The color curve indicates the feasibility function $\mathcal{F}((\gamma_0, S_0, \alpha_0 = 0), G, 0)$ for G_1, which has a connectance $\kappa_G = 0.18$ and ecological overlap $\eta_G = 0.2$ (left) and G_2 with $\kappa_G = 0.28$ and $\eta_G = 0.4$(right). We observe a steep descent which marks a very clear transition from a totally feasible regime to a totally unfeasible regime, which allows us to precisely get the boundary of $\mathcal{F}_1^{G,0}$. The dashed lines indicate the theoretical predictions.

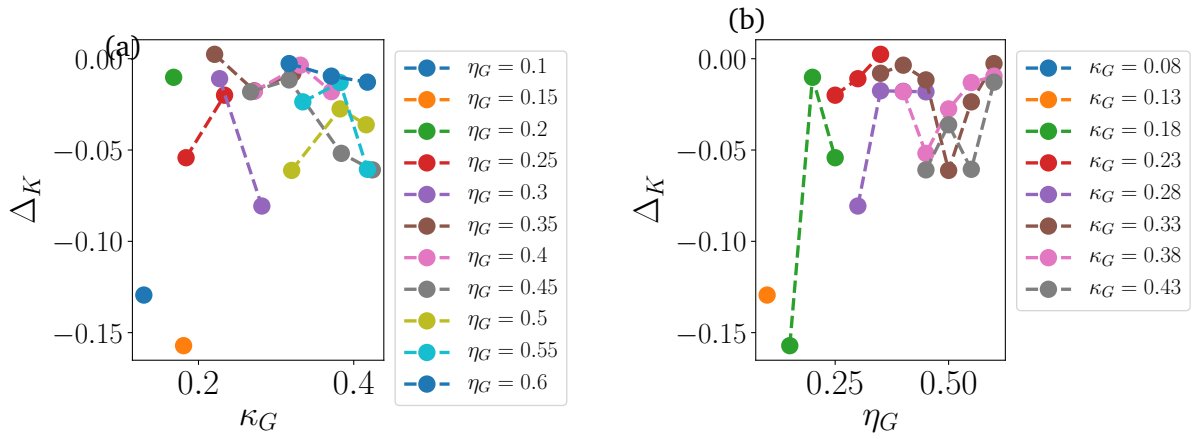

Figure 3.1.2: Relative error in the determination of the boundary of $\mathcal{F}_1^{G,0}$ (a) varying connectance at fixed ecological overlap and (b) varying ecological overlap at fixed connectance. The theoretical prediction tends to overestimate the measured value. The larger the ecological overlap or connectance, the better the estimate.

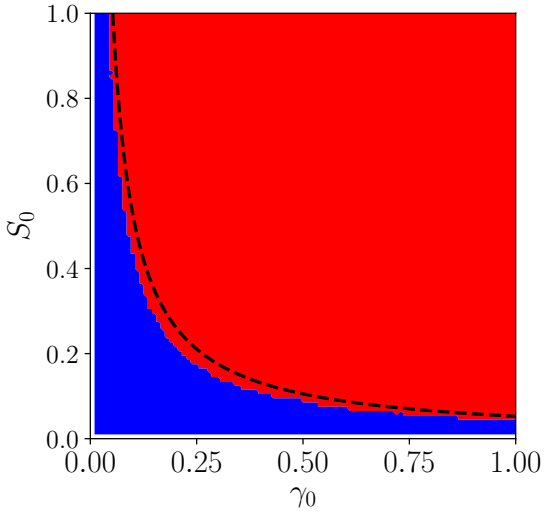

Figure 3.1.3: Plot of the common feasibility region. The blue area indicates the common feasibility volume, computed numerically, while the dashed line shows the analytical prediction. Although the match is not as good as before, the relative error is only of the order of 20%. The red part is the area where not all matrices are fully feasible.

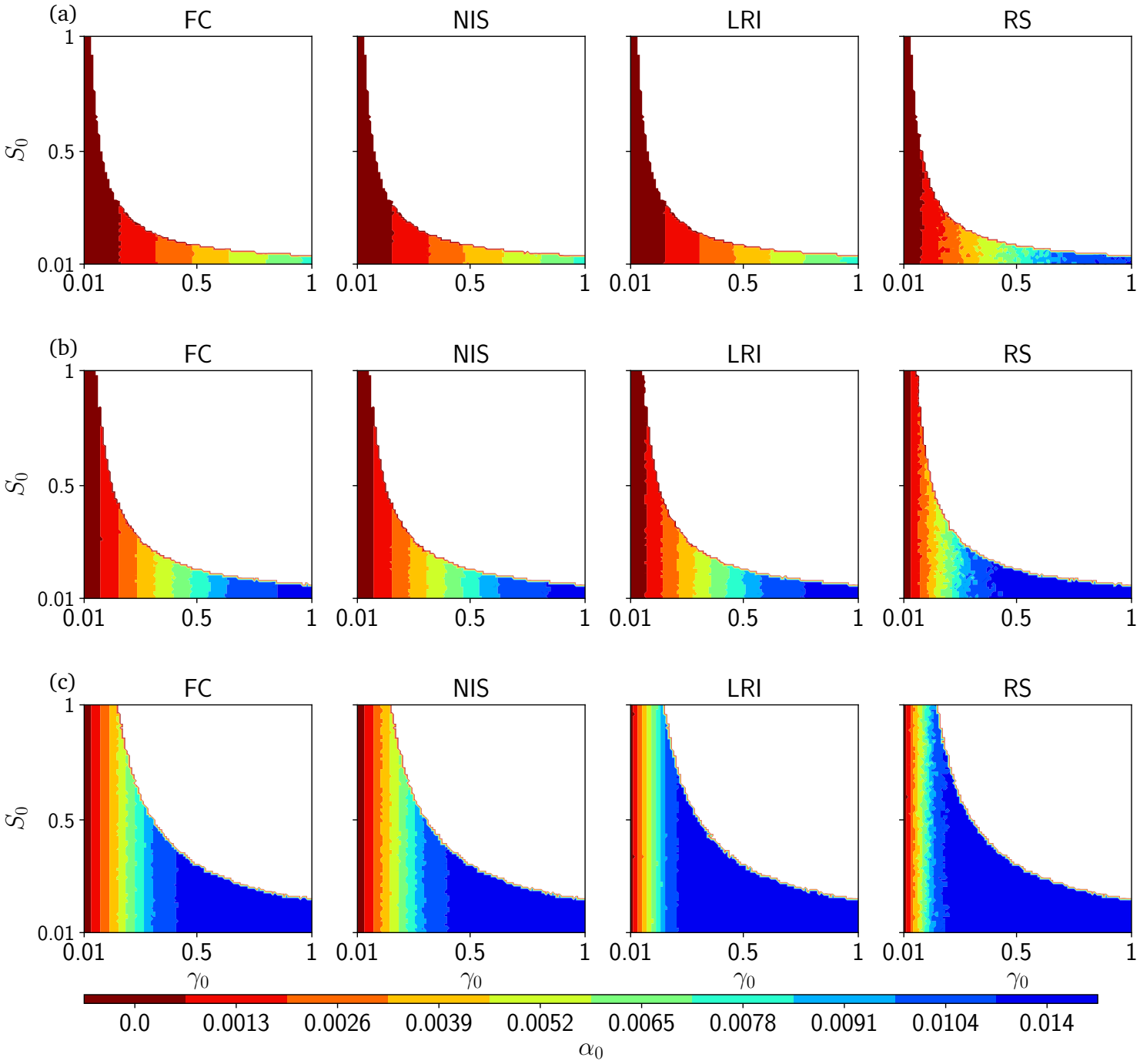

Figure 3.1.4: Fully feasible region in the $(\gamma_0, S_0) \in [0,1] \times [0,1]$ unit square as a function of syntrophy for different consumption matrices G: (a) $\eta_G = 0.6$, $\kappa_G = 0.32$, (b) $\eta_G = 0.35$, $\kappa_G = 0.22$ and (c) $\eta_G = 0.15$, $\kappa_G = 0.18$. The white zone corresponds to points that are never fully feasible. The colour of a given point tells until which syntrophy that point is fully feasible, e.g. a light blue point is fully feasible for $0 \leqslant \alpha_0 \leqslant 9.1 \times 10^{-3}$. The size of the feasibility regions depend heavily on the topology of the matrix, which makes the problem far from trivial.

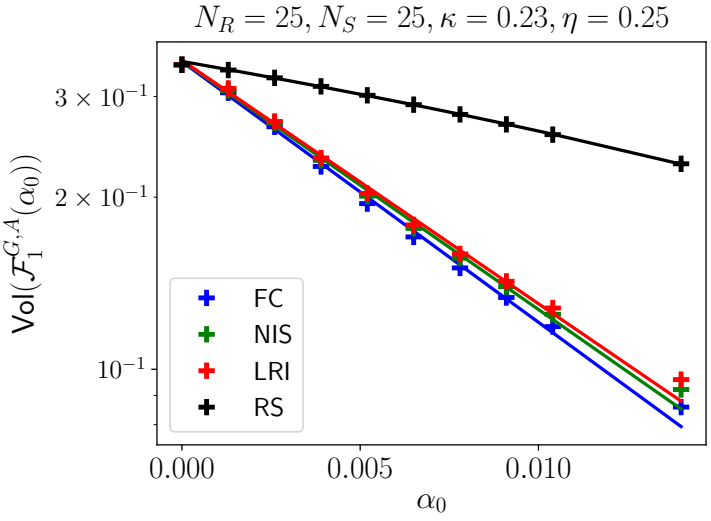

Figure 3.1.5: Decay of the volume of the fully feasible region $\mathcal{F}_1^{G,A}(\alpha_0)$ for a matrix consumption G with ecological overlap $\eta_G = 0.25$ and connectance $\kappa_G = 0.23$ on a logarithmic scale. The solid lines represent the exponential fit explained in the main text. The four different colors represent the four different structures considered for the syntrophy matrix. The decay of $\mathrm{Vol}(\mathcal{F}_1^{G,A}(\alpha_0))$ seems well approximated by an exponential decay. A random syntrophy matrix (RS scenario) allows for a larger feasibility volume.

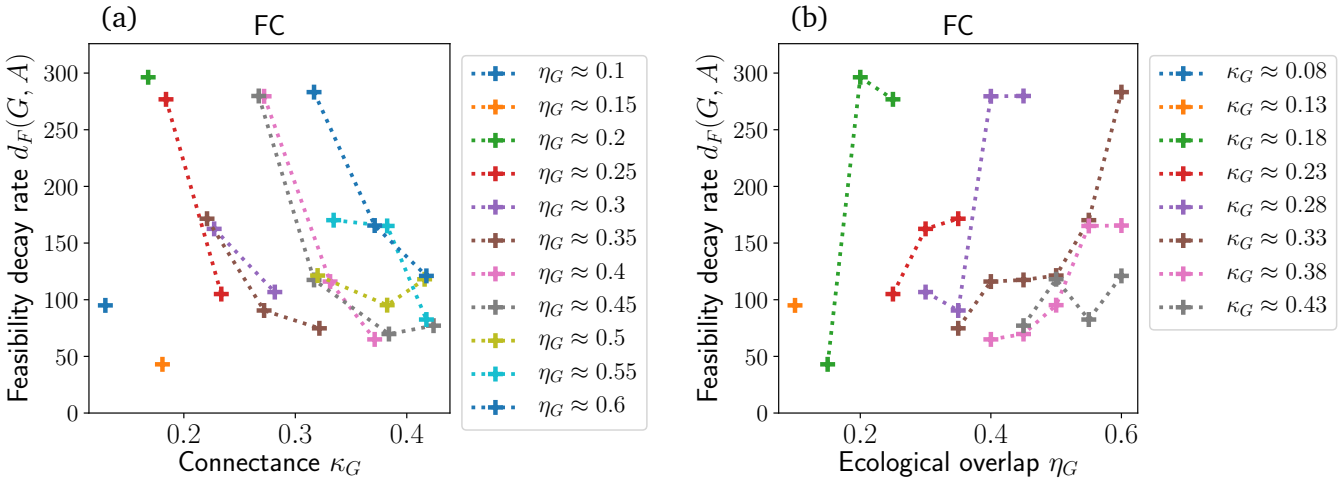

Figure 3.1.6: Feasibility decay rate $d_F(G, A)$ for A fully connected and $(G, A) \in S_{25}$. (a) d_F as a function of the connectance of G for different fixed ecological overlaps and (b) d_F as a function of the ecological overlap η_G for fixed different connectances. A strong trend may be seen: at fixed ecological overlap, d_F decreases with connectance and at fixed connectance it increases with ecological overlap. Since a small d_F allows to sustain a larger syntrophy, microbial communities where syntrophic interactions play a large role will tend to have a high connectance of the consumption matrix and a low ecological overlap.

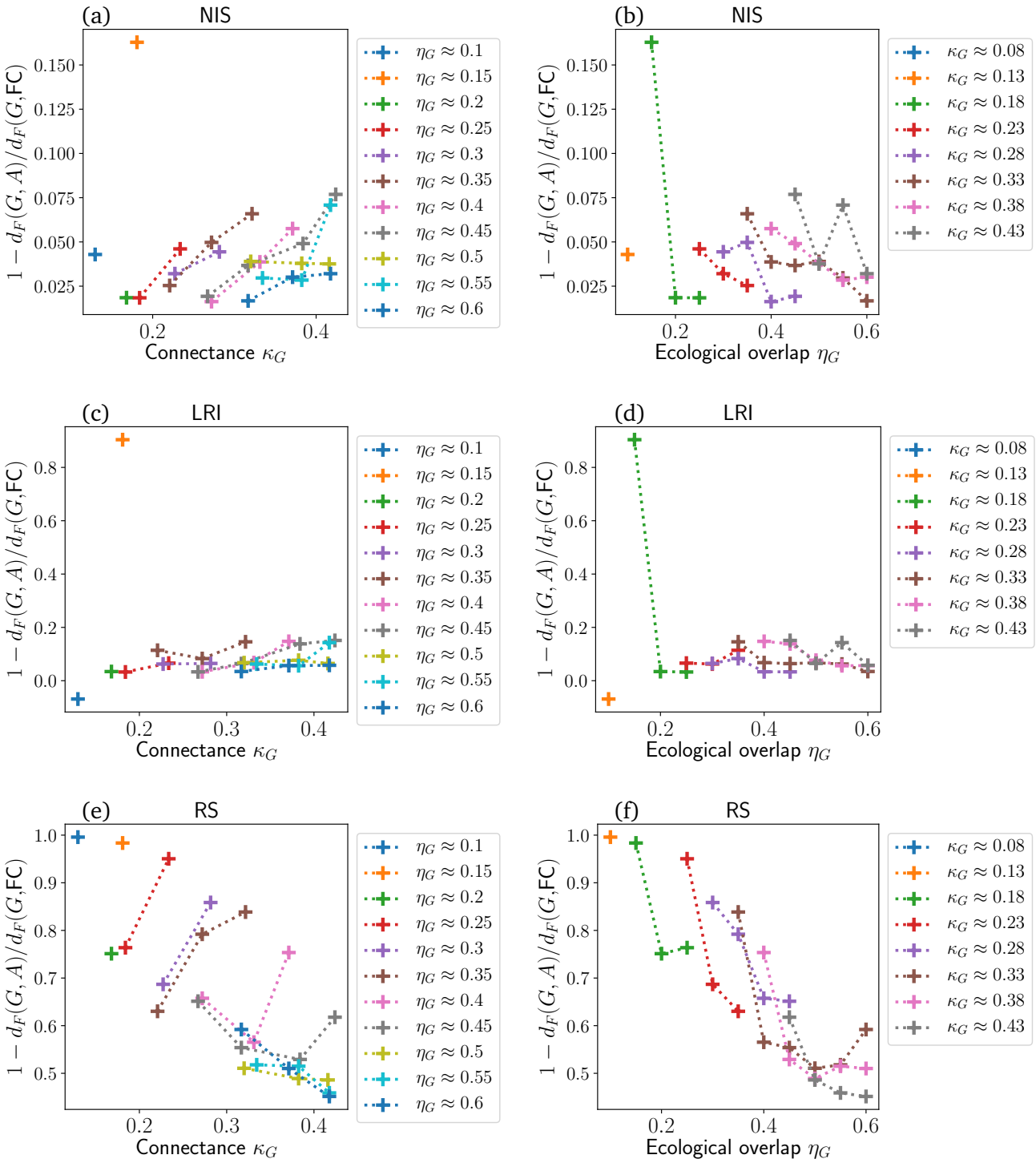

Figure 3.1.7: Relative difference of the feasibility decay rate for the considered A scenario compared to the FC null case (Fig.3.1.6) for $N_R = 25$ and $N_S = 25$. Plots on the first column (a)-(c)-(e) show how that quantity changes with connectance for a given ecological overlap, while plot on the second column (b)-(d)-(f) show how it evolves when ecological overlap is changed and connectance is kept fixed. Different structures of the A matrix are considered: (a)-(b) NIS, (c)-(d) LRI (e)-(f) RS. A positive y-coordinate means that for the feasibility decay rate of the current syntrophy scenario is smaller than for the FC case, i.e. the system sustains syntrophy better with the considered A-structure compared to fully connected. Apart from a few marginal exceptions, the FC scenario is always outperformed by the other scenarios, especially the random structure (RS) case.

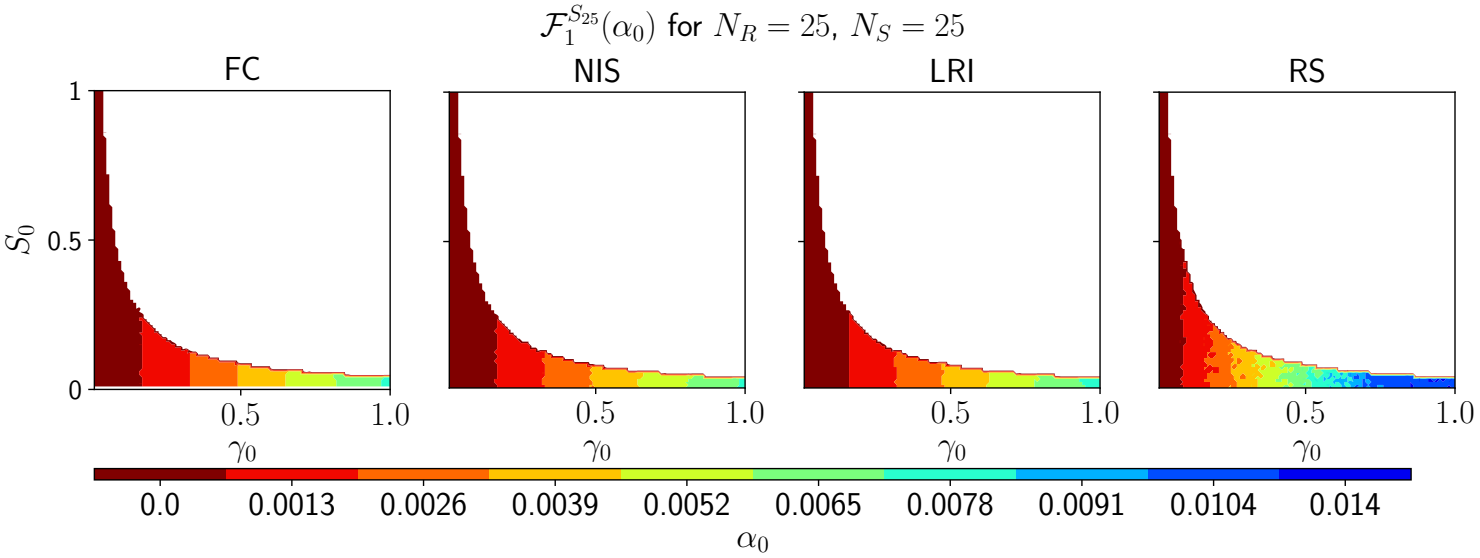

Figure 3.1.8: Surface plot of the fully feasible volume $\mathcal{F}_1^{S_{25}}(\alpha_0)$. The color bar on the side indicates the value of α_0 to which the surface corresponds. The white part of the plot corresponds to points that never are fully feasible. Note that even though it is not very clear on the figure $\mathcal{F}_1^{S_{25}}(\alpha_0^+) \subset \mathcal{F}_1^{S_{25}}(\alpha_0^-)\ \forall \alpha_0^+ > \alpha_0^-$, i.e. the common fully feasible region of higher syntrophy is included in the one of lower syntrophy. The different subplots correspond to the different structures of the syntrophy matrix.

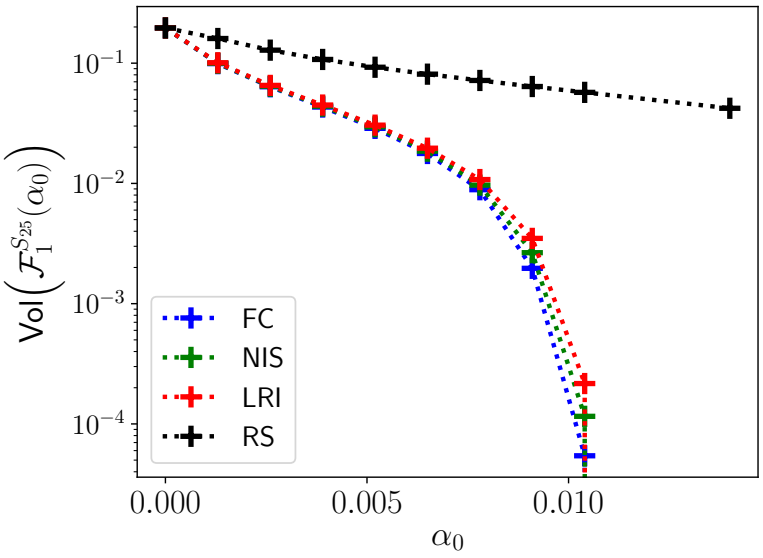

Figure 3.1.9: Volume of the common feasibility region $\mathcal{F}_1^{S_{25}}(\alpha_0)$ as a function of syntrophic interaction strength α_0 (plotted on logarithmic scale). While the FC, NIS and LRI cases offer similar results, the RS scenario outperforms all of them. An exponential fit (Eq.3.4) allows to measure a global d_F for each of the four scenarios. The global decay feasibility rate of the RS scenario is 2.5 times smaller than the others.

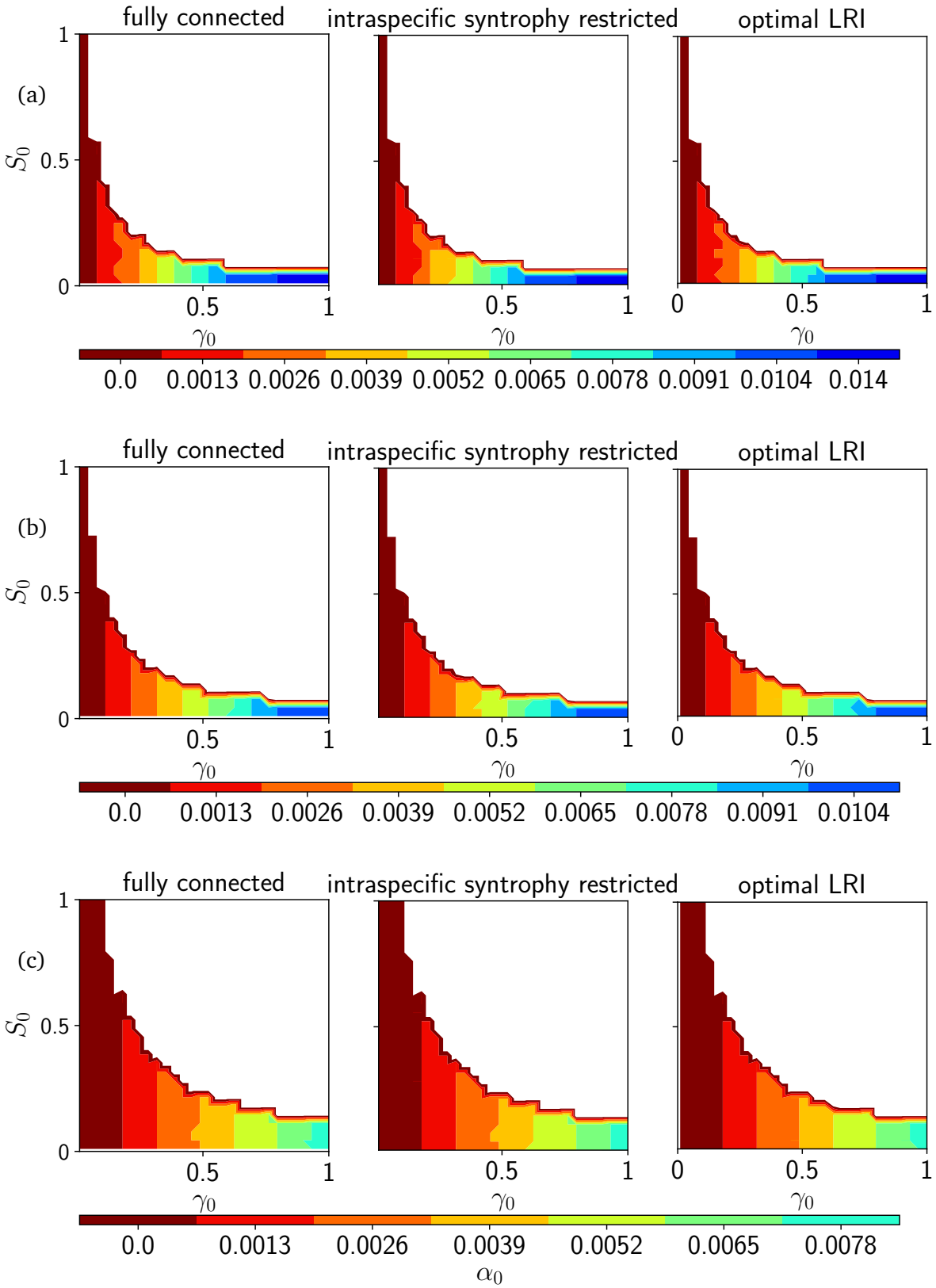

Figure 3.1.10: Surface colour plot of the fully feasible region $\mathcal{F}_1^{G,A}(\alpha_0)$ as a function of the syntrophy α_0 for the case $N_R = 50$, with different structures of A: fully connected (left column), no intraspecific syntrophy (middle) and LRI matrix (right). The rows correspond to different choices of the consumption matrix G: (a) $\eta_G = 0.6$ and $\kappa_G = 0.33$, (b) $\eta_G = 0.35$ and $\kappa_G = 0.23$, (c) $\eta_G = 0.15$ and $\kappa_G = 0.12$. These are matrices with similar properties than Fig.3.1.4, except that the number of resources is here doubled. This affects $\mathcal{F}_1^{G,A}(\alpha_0)$ quite drastically.

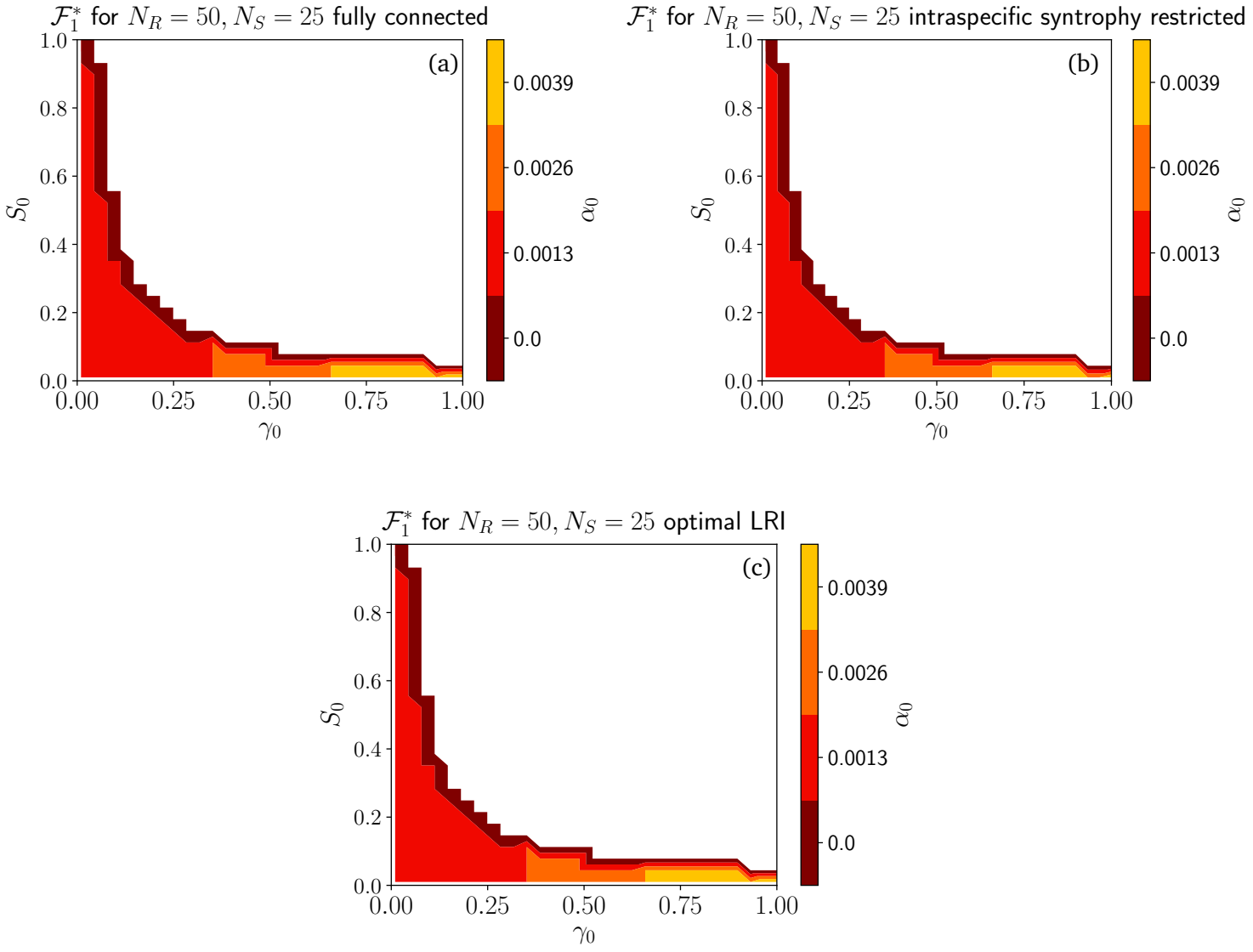

Figure 3.1.11: Common feasibility region $\mathcal{F}_1^{S_M}(\alpha_0)$ for $N_R = 50$ and $N_S = 25$, to compare with 3.1.8. We considered different structures of the syntrophy matrix: (a) fully connected, (b) intraspecific syntrophy restricted and (c) LRI matrix. As the number of resources increases, the feasibility volume for a given α_0 decreases.

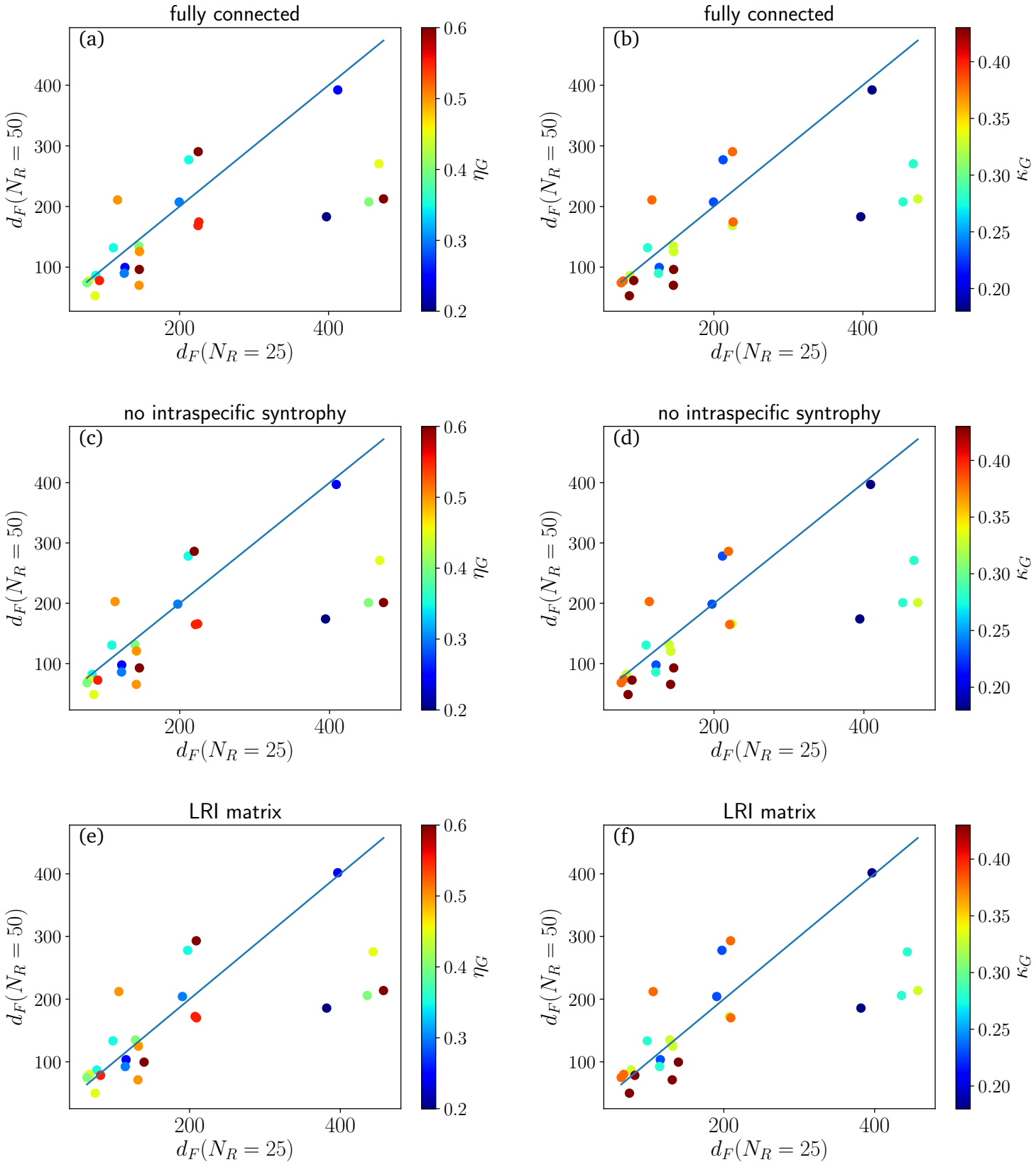

Figure 3.1.12: Comparison of the feasibility decay rates obtained for matrices with same ecological and connectance, at $N_R = 25$ and $N_R = 50$. The color of the points indicates their ecological overlap (left) or connectance (right). The straight line represents equal $d_F(N_R = 50)$ and $d_F(N_R = 25)$. Different structures for the syntrophy matrix A are considered: (a) fully connected, (b) no intraspecific syntrophy and (c) LRI matrix. No clear tendency on what effect increasing resources will have can be drawn.

3.2 Dynamical stability

TO DO: put small introduction here

3.2.1 LRI regime – Outcome of the MCMC algorithm

Methods 2.2.4 explains how we designed an algorithm whose goal is to find, for a given consumption matrix G, the syntrophy matrix A that should bring us closer to the satisfaction of Eq.(2.76), which we know is a dynamically stable regime. The shape of the A-matrix that is the outcome of the algorithm minimizes the energy $E(G, A)$ (Eq.2.93). Namely, A should have the following properties:

- The sum of the diagonal elements of AG is minimized. Because $(AG)_{\mu\mu} = \sum_{i=1}^{N_S} A_{\mu i} G_{i\mu}$ corresponds to the number of species that both consume and release resource μ, we expect the algorithm to yield A-matrices that minimize intraspecific syntrophy.
- The sum of the off-diagonal elements of $\left|\alpha_0 AG - \gamma_0 R_0 G^T G\right|$ is minimized. This means that outside the diagonal, we should have $\frac{\alpha_0}{\gamma_0 R_0} AG \approx G^T G$. A direct ecological interpretation is harder to draw. For a couple of different resources (μ, ν), $(AG)_{\mu\nu} = \sum_{i=1}^{N_S} A_{\mu i} G_{i\nu}$ is the number of species that at the same time consume resource ν and release resource μ and $(G^T G)_{\mu\nu} = \sum_{i=1}^{N_S} G_{i\mu} G_{i\nu}$ is the number of consumers that eat both ν and μ. **TO DO: add something here, transition is a bit rough**

So intuitively the LRI MCMC algorithm should give us a syntrophy matrix that both limits intraspecific syntrophy and such that for every couple of different resources (μ, ν), the number of consumers that eat both μ and ν is proportional to the number of consumers that eat μ and release ν. The proportionality constant, which is the same for all (μ, ν) is equal to the ratio of the syntrophy and consumption interactions. Since the connectance and the dimensions of A are fixed, the number of links of A is already decided and the algorithm simply determines how to optimally distribute them. Figure 3.2.1 shows that typically the algorithm will put links in a cell (μ, i) if (i, μ) is zero, meaning that not only intraspecific syntrophy tries to be avoided but also species that consume a lot of resources will tend to release few of them and vice-versa.

Figure 3.2.2 shows that indeed we obtain for a given G matrix a syntrophy matrix A such that the two requirements above are best satisfied. Note that the algorithm works better for matrices with a low connectance. **TO DO: explain why** It is worth noticing that this procedure produces highly nested syntrophy matrices (Fig.3.2.3) where only a few species produce most of the syntrophic flow. The obtained matrices have an even larger nestedness if we increase the number of resources. **TO DO : check if can speak of syntrophic overlap for nestedness of syntrophy matrix**

3.2.2 Typical dynamical stability

We observed in Results 3.1.1 that, for all $(G, A) \in S_{25}$, at fixed α_0 the metaparameters feasibility function $\mathcal{F}(m, G, A)$ has a typically sharp transition from fully feasible $(\mathcal{F} = 1)$

to fully unfeasible ($\mathcal{F} = 0$) regimes in the (γ_0, S_0) plane. Figure 3.2.4 shows that a similar although more complicated behaviour is observed in the case of the dynamical stability function $\mathcal{D}_L\left((\gamma_0, S_0, \alpha_0), G, A\right)$. On one hand, the (γ_0, S_0) plane is split in two distinct zones, which are also separated by a very narrow boundary. The first zone is characterised by complete dynamical instability, *i.e.* $\mathcal{D}_L = 0$. On the other hand, the second zone is not described by full dynamical stability, but rather *almost* full dynamical stability: $\mathcal{D}_L$ is very close to but not always exactly equal to 1. The consequence is that the fully dynamically stable region $\mathcal{D}_{L,1}^{G,A}$ will be very patchy (see below for a longer discussion on this).

That patchiness could come from purely numerical effects: $\mathcal{D}_L$ is estimated by generating N_{sys} parameters sets and counting the proportion that is dynamically stable, which inevitably leads to an uncertainty on $\mathcal{D}_L$ that could explain the patchiness. Or it could come from a genuine complicated topology of $\mathcal{D}_{L,1}^{G,A}$. In a future project, increasing N_{sys} would allow to reduce the relative uncertainty on $\mathcal{D}_L$ and truly discover the origin of this interesting phenomenon.

3.2.3 Fully dynamically stable region

The same way we studied the fully feasible volume $\mathcal{F}_1^{G,A}(\alpha_0)$, we investigate now the behaviour of its special subset, the locally fully dynamically stable region $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$, which is defined[27] as

$$\mathcal{D}_{L,1}^{G,A}(\alpha_0) \equiv \left\{(\gamma_0, S_0) : (\gamma_0, S_0, \alpha_0) \in \mathcal{D}_{L,1}^{G,A}\right\}. \tag{3.6}$$

Intuitively, $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$ corresponds to the set of all (γ_0, S_0) such that $\mathcal{A}\left((\gamma_0, S_0, \alpha_0), G, A\right)$ is a feasible, locally dynamically stable parameters set with probability 1. Since we require $\mathcal{A}\left((\gamma_0, S_0, \alpha_0), G, A\right)$ to be feasible, it is clear that $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$ is indeed a subset of $\mathcal{F}_1^{G,A}(\alpha_0)$.

As explained above, $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$ is geometrically more complex than $\mathcal{F}_1^{G,A}(\alpha_0)$ (Figure 3.2.5) . It may sometimes have holes, even without syntrophy, and sometimes not, even for matrices that are topologically very close. Compare for instance Fig.3.2.5a with Fig.3.2.5c, these two networks have the same ecological overlap, but even though their connectance is very similar, their fully locally dynamically stable regions have a very different shape: one of them can sustain only a tiny bit of syntrophy before becoming dynamically unstable (Fig.3.2.5a) while the second can endure basically any feasible syntrophic interaction (Fig.3.2.5c). A general trend is hence harder to find but it seems that points with a larger γ_0 are in general more dynamically stable (more on this below).

We required that a system had to be feasible in order to be dynamically stable, which is the mathematical equivalent of $\mathcal{D}_{L,1}^{G,A}(\alpha_0) \subset \mathcal{F}_1^{G,A}(\alpha_0)$, *i.e.* local dynamical stability implies feasibility. We may also ask the reverse question, does feasibility imply local dynamical stability? The answer to this is, again, unsurpisingly, "it depends on the matrix", as shows Figure 3.2.5. For instance, for G with $\kappa_G = 0.13$ and $\eta_G = 0.1$, we have $\text{Vol}\left(\mathcal{D}_{L,1}^{G,A}(\alpha_0)\right) < \text{Vol}\left(\mathcal{F}_{L,1}^{G,A}(\alpha_0)\right) \forall \alpha_0$, which means that for this consumption matrix feasibility does not imply stability. The fully connected case gives a larger dynamically stable volume than the regime without intraspecific syntrophy which is itself better than the LRI regime. This hints that the LRI regime, despite what it was designed for, apparently does

[27] A formal definition of the fully dynamically stable region $\mathcal{D}_{L,1}^{G,A}$ is provided in Methods 2.2.1.

not give better results than other structures of A. On the contrary, for G with $\kappa_G = 0.32$ and $\eta_G = 0.6$, both volumes are equal at every syntrophy that is feasible, for the three structures of A considered, which shows that for this specific matrix, feasibility implies local dynamical stability.

A good way to measure how systems react to syntrophy is to compute the *critical locally dynamically stable syntrophy* $\alpha_0^D(G, A)$ (see Methods **??**). An easy way this can be done is by getting some points of the volume of $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$ curve and finding its intercept to zero. Figure 3.2.7 shows the typical shrinkage of the fully locally dynamically stable volume.

To find $\alpha_0^D(G, A)$ for each G, we fit with a linear function the last four points of the curve corresponding to Fig.3.2.7 and find its intercept to zero $\alpha_0^D(G, A)$. Figure 3.2.8 shows a very interesting and clear behaviour[28]: for a given connectance of the consumption matrix, systems that can sustain the largest syntrophy have a small ecological overlap. And for a given ecological overlap, systems with a larger connectance will stay stable longer under the action of syntrophy. In the end, optimal systems have a small ecological overlap and a large connectance: many resources are eaten by the consumers, but they do not share them.

Finally, one can take a look at the *common fully locally dynamically stable region*, which is the intersection of the $\mathcal{D}_{L,1}^{G,A}(\alpha_0) \forall (G, A) \in S_M$ (Figure 3.2.9). Because of the fractured and heterogenous nature of each $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$, we observe a very fractured and small common fully locally dynamically stable region, which is the same for all structures of A considered. It has a non-zero volume for $\alpha_0 = 0$, but for the next point investigated $\alpha_0 = 1.3 \times 10^{-3}$, no point is fully locally dynamically stable for every matrix considered, which means that the critical common syntrophy is smaller than this.

3.2.4 Largest eigenvalue of the jacobian

Equation (**??**) from Methods **??** gives a relationship that the metaparameters should approximately follow in order to give rise to locally dynamically stable systems. Although strictly speaking it is only valid for the case where both G and A are fully connected, we expect it to work as well when G and A are *not too far away* from the fully connected case. It tells us that in order to get more local dynamically stable systems you should:

- decrease N_S, l_0 **WEIRD RESULT: would expect that increasing l0 would make systems more dynamically stable (observed in simulations I think)** or α_0,
- if $\alpha_0 - \gamma_0 R_0 < 0$, increase N_R, σ_0 and γ_0,
- be careful in how you handle S_0: increasing S_0 reduces the l_0^2/S_0 term but increases the $N_S^2 \alpha_0^2 S_0$ term. It is very easy to show (Appendix 5.1.2) that if $S_0 > l_0/(N_S \alpha_0)$ it should be decreased, and otherwise it should be increased until it reaches $l_0/(N_S \alpha_0)$.

Combining these with the feasibility conditions Eq.(**??**) we expect that – for all other metaparameters fixed – systems get more and more locally dynamically stable as γ_0 is

[28] Those results have to be taken with a grain of salt because we did not check whether $\alpha_0^D(G, A)$ was feasible, *i.e.* the actual critical locally dynamically stable syntrophy is the minimum between the value measured in Fig.3.2.8 and the largest feasible α_0 for that couple (G, A).

increased and S_0 is the largest possible. In short, points at the upper border of $\mathcal{D}_{L,1}^{G,A}$ should have a lower and lower $\mathrm{Re}\,(\lambda_1)$ as γ_0 increases. Figure 3.2.10 shows that indeed this trend is followed. This tells us that if we keep the consumption flux $N_S\gamma_0 S_0$ constant, increasing γ_0 (and hence decreasing S_0) will give rise to more stable systems. Notice that contrarily to the prediction made above, increasing α_0 does not decrease stability but increases the maximal $|\mathrm{Re}\,(\lambda_1)|$ observed as shows Fig.3.2.11a. This is coupled with the shrinkage of the fully locally dynamically stable volume seen on Fig.3.2.11b. This means that overall increasing syntrophy makes the system *more stable* but at *fewer points*. This hints that systems in a high syntrophic regime, where consumers produce a lot of resources, should be very fine-tuned and occur for very specific consumption strength and average abundance of consumers.

3.2.5 The influence of the matrix dimension

As said above, because of Eq.(**??**), we expect stability to increase when the number of resources is increased for a fixed number of consumers. The following subsection shows what happens when the number of resources is doubled $N_R = 25 \to N_R = 50$ and every other metaparameter, as well as the number of consumers, keeps the same value as before.

Figure 3.2.12 shows that the effect of adding resources can be quite dramatic on the stability of the system. For that specific matrix for instance, adding resources allowed for a way larger $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$ at each α_0. It even allows full dynamical stability in the feasibility region for larger syntrophic interactions than before. Even though more syntrophy can be sustained, it seems that the volume of $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$ is smaller at $N_R = 50$ than at $N_R = 25$ (compare for instance Fig.3.2.13b with Fig.3.2.11b). This is compensated by the fact that way larger eigenvalues are observed at $N_R = 50$: although there are (a bit) fewer equilibrium points, these are more stable (compare Fig.3.2.13a and Fig.3.2.11a). This is a trend that we believe holds for all the matrices considered but a more thorough investigation should be conducted before claiming those results to be absolutely true.

Since matrices are individually more locally dynamically stable, the *common locally dynamically stable* region is larger with $N_R = 50$ resources than with $N_R = 25$. The difference between Figure 3.2.9 and Figure 3.2.14 is very striking: when the number of resources is increased not only is $\mathcal{D}_{L,1}^{S_M}(\alpha_0)$ larger, it also can overall bear a larger syntrophy. Remember that for $N_R = 25$, the critical common syntrophy was between 0 and 1.3×10^{-3}, and for $N_R = 50$, it is way larger, between 3.9×10^{-3} and 5.2×10^{-3}. **Is there anything more to add? I have some plots but I am not sure if they are the most relevant**

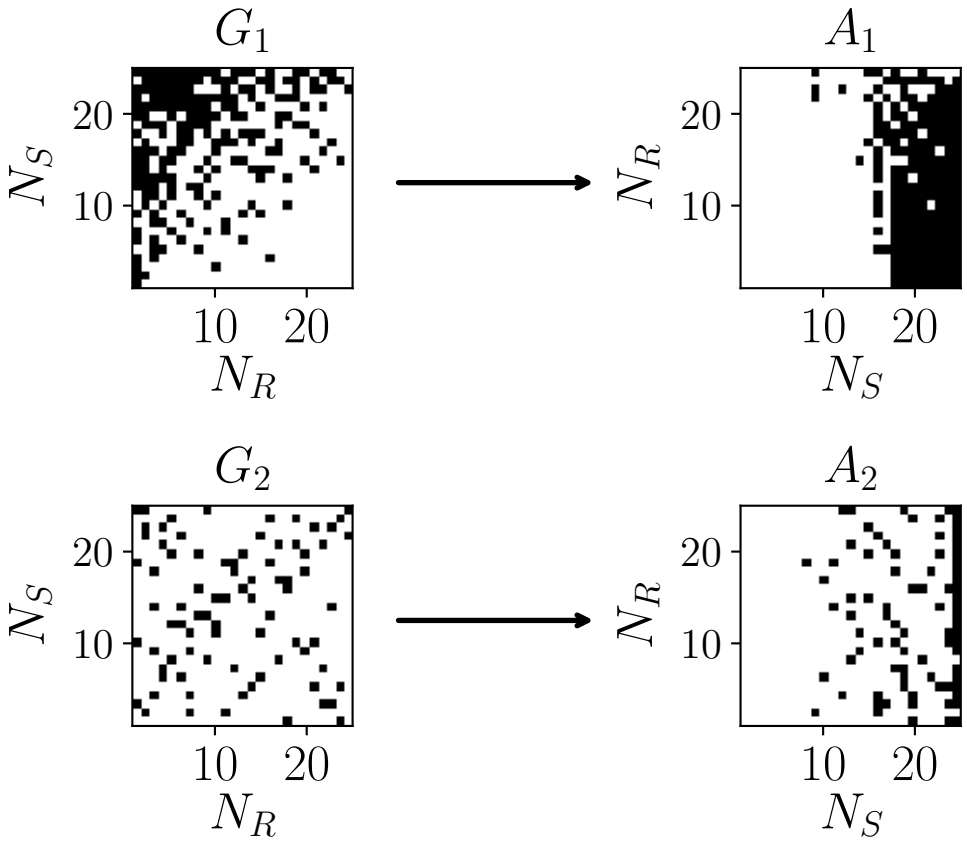

Figure 3.2.1: Typical shape of the consumption G_i and syntrophy A_i matrices. The white cells symbolize a zero matrix element and the black cells, a one. A_i here is the outcome of the LRI MC algorithm described in Methods 2.2.4. The first row has a consumption matrix with $\eta_1 = 0.6$ and $\kappa_1 = 0.32$, the LRI MC solver gives rise to a syntrophy matrix with same connectance and ecological overlap ≈ 0.85. The second row has G_2 with $\eta_2 = 0.1$ and $\kappa_2 = 0.13$ and the corresponding syntrophy matrix A_2 has ecological overlap ~ 0.42. We observe that under this optimisation, species that consume few resources end up releasing many and the other way around.

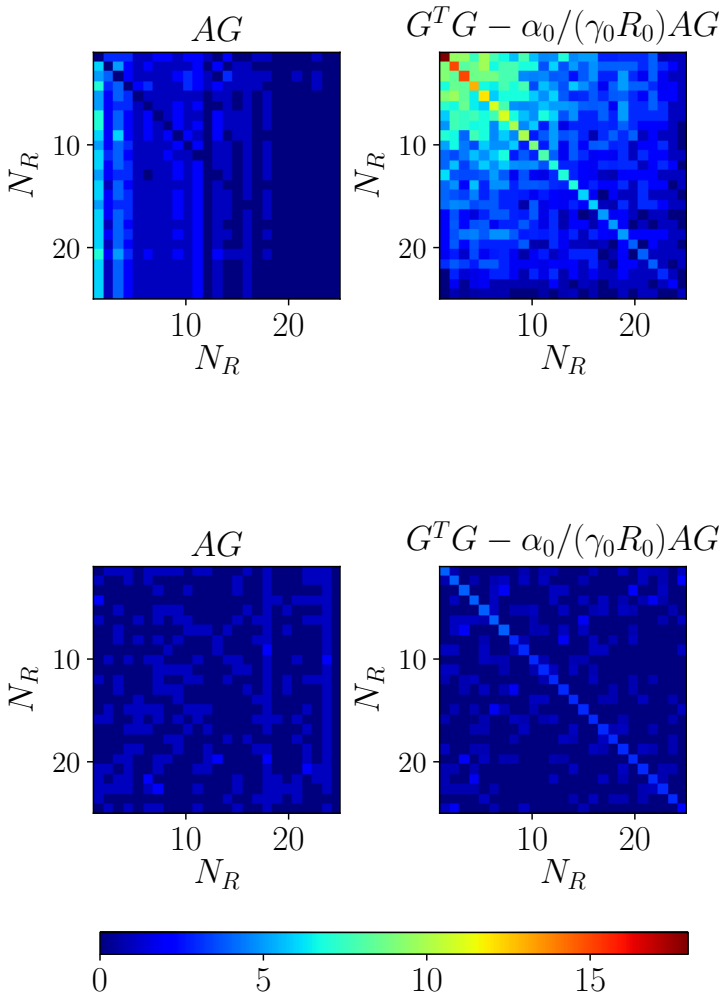

Figure 3.2.2: Plotting of AG and $G^TG - \alpha_0/(\gamma_0 R_0)AG$. The A and G matrices of the first and second rows correspond to the respective A and G of Figure 3.2.1. As expected, we obtain an A such that intraspecific coprophagy is limited (the diagonal of AG is roughly zero) and, outside the diagonal, $G^TG - \alpha_0/(\gamma_0 R_0)AG \approx 0$. Both relations are better satisfied for consumption (and hence syntrophy) matrices with a low connectance.

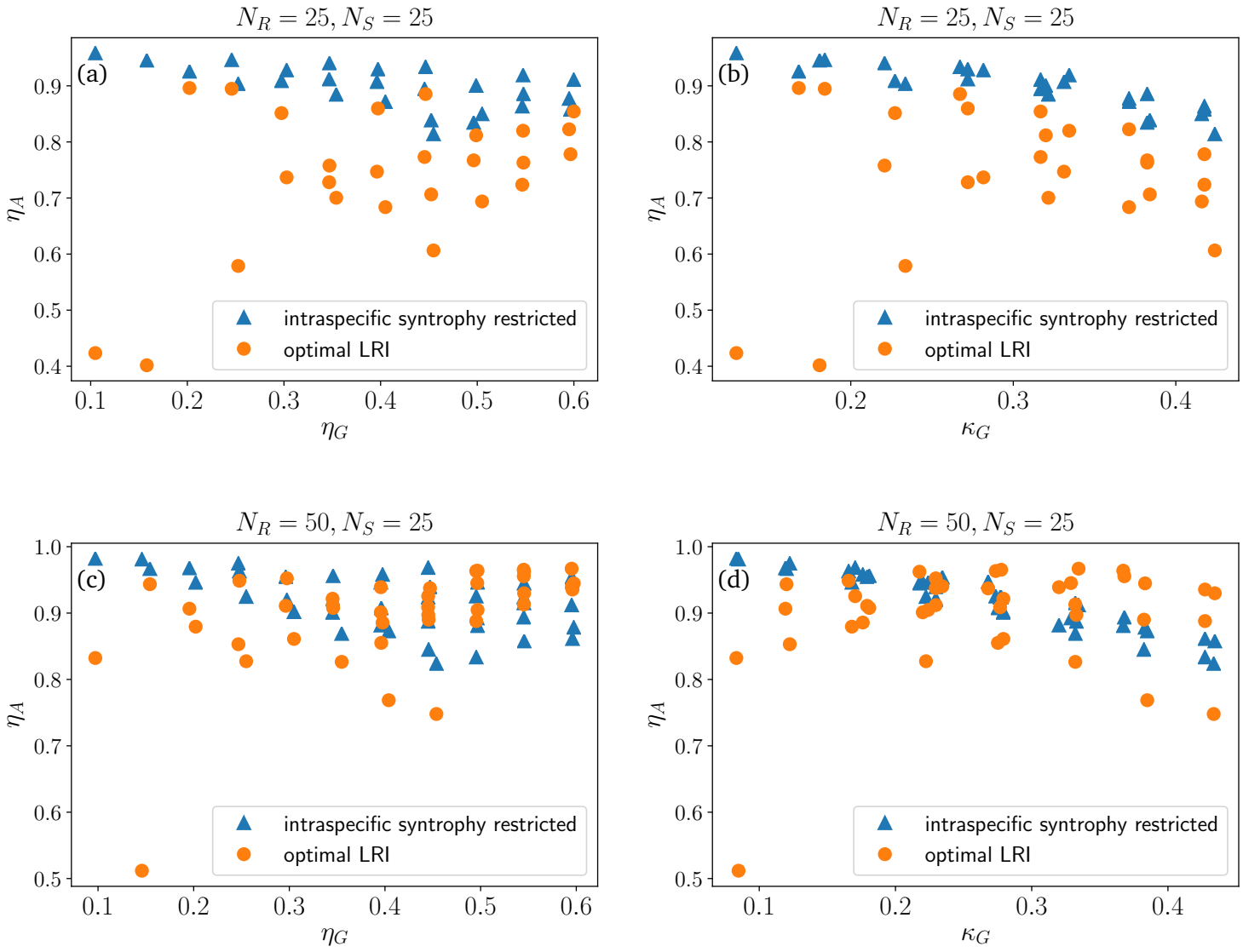

Figure 3.2.3: Properties of the syntrophy matrix against the consumption matrix. (a)-(c) Ecological overlap of A as a function of the ecological overlap of G for $N_S = 25$ and $N_R = 25$ (a) or $N_R = 50$ (c). (b)-(d) Ecological overlap of A as a function of the connectance of G for $N_S = 25$ and $N_R = 25$ (b) or $N_R = 50$ (d). The nestedness of the "intraspecific syntrophy restricted" is also plotted as a matter of comparison. As η_G or κ_G increase, the two results will without surprise give matrices with similar properties. ***Explain this?***

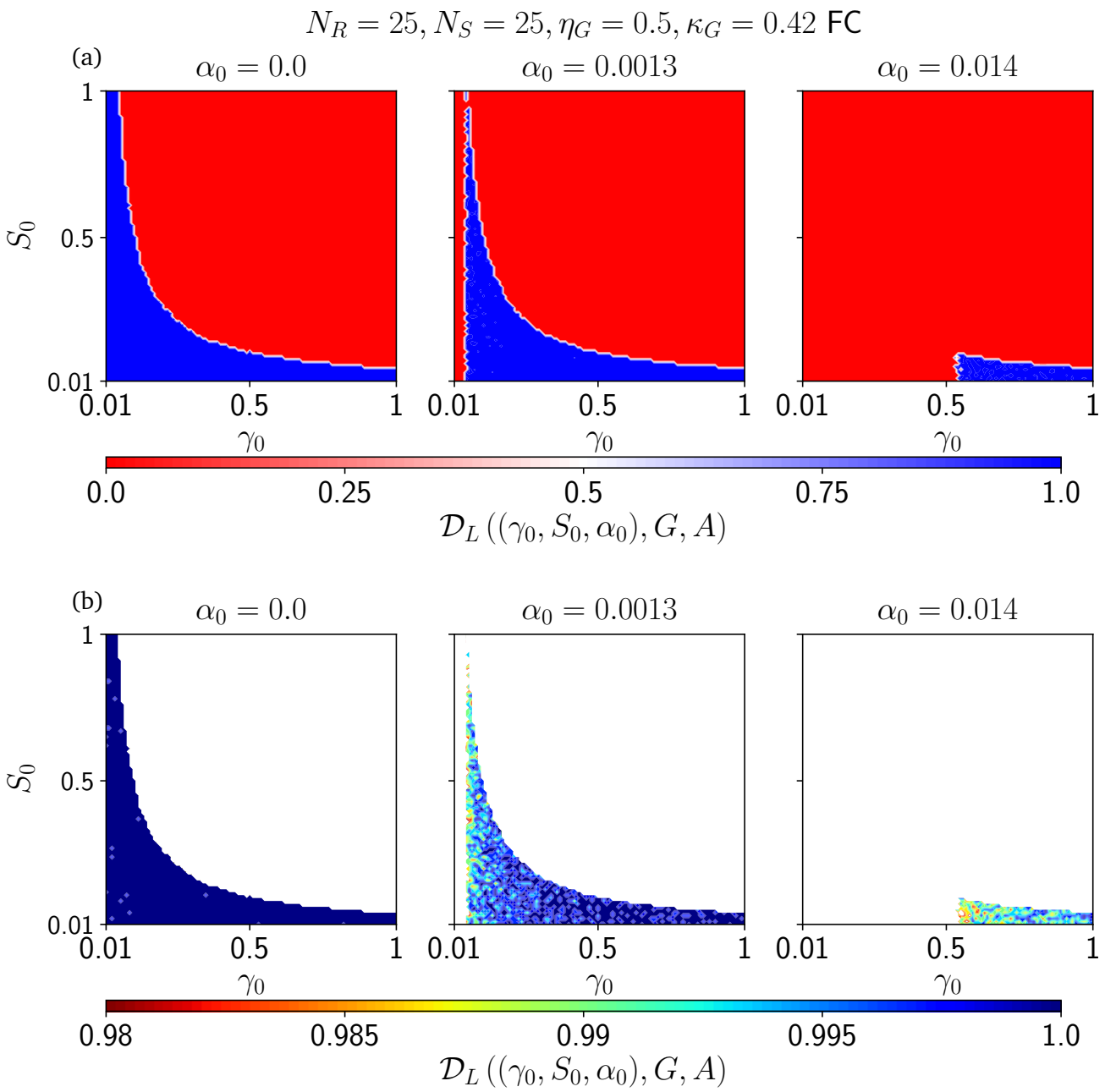

Figure 3.2.4: Typical color plot local dynamical stability metaparameters function $\mathcal{D}_L$. for microbial communities with $\eta_G = 0.5$ and $\kappa_G = 0.42$. The color bar indicates the value of $\mathcal{D}_L((\gamma_0, S_0, \alpha_0), G, A)$ with A fully connected. The plane is divided in two zones, one where $\mathcal{D}_L = 0$ and another where $\mathcal{D}_L \approx 1$ (the red and blue regions, respectively, in (a)). Upon further notice (b), it turns out the $\mathcal{D}_L \approx 1$ region is very patchy: we observe many $(\gamma_0, S_0, \alpha_0)$ configurations for which $\mathcal{D}_L$ is very close but not exactly equal to 1. These points are almost fully dynamically stable.

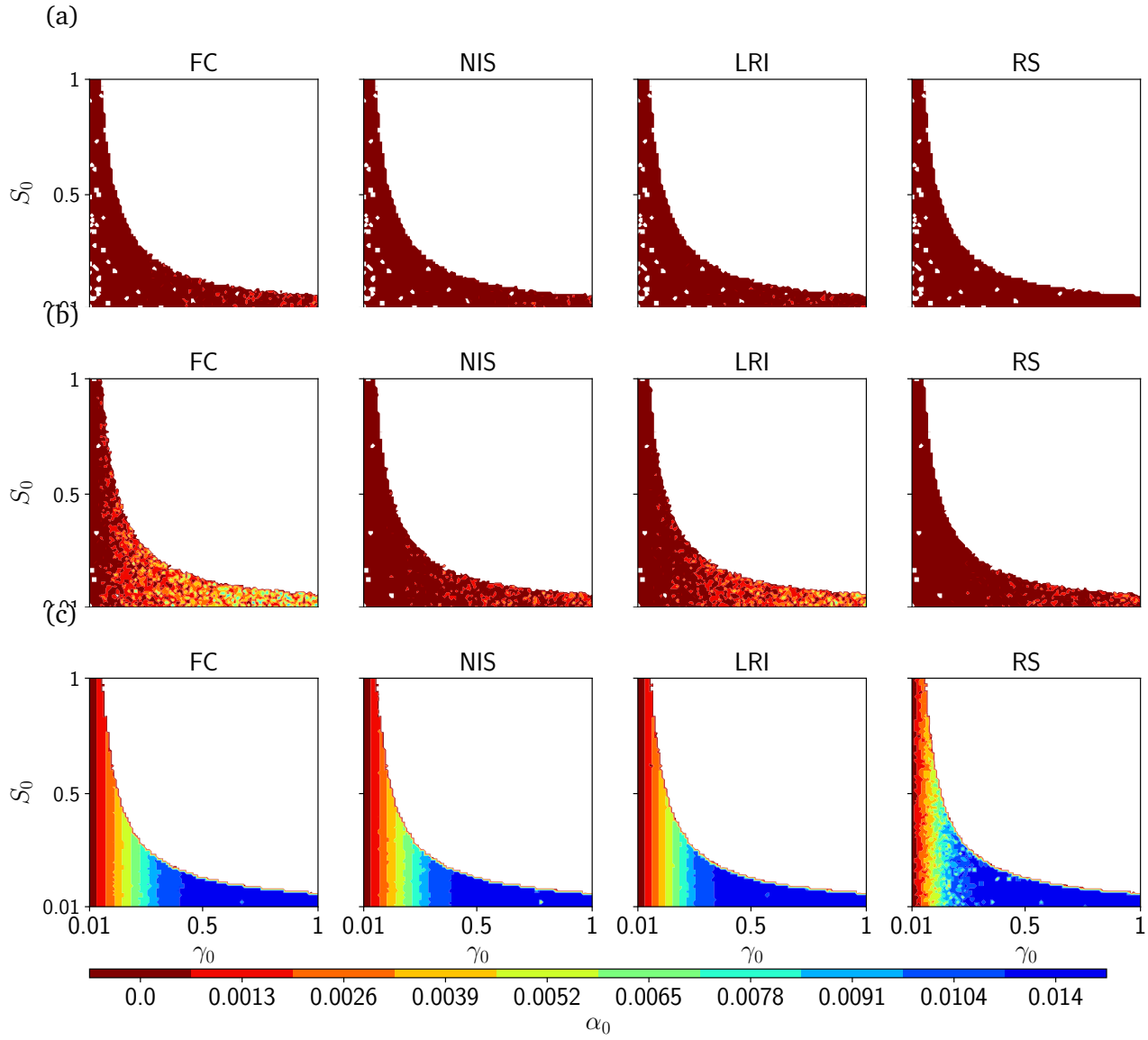

Figure 3.2.5: Locally fully dynamically stable region $\mathcal{D}_{L,1}^{G,A}$ as a function of syntrophy for different matrices G. The white zone corresponds to points that are never fully locally dynamically stable. The colour of a given point tells until which syntrophy that point is fully locally dynamically stable, e.g. a green point is fully locally dynamically stable for $0 \leqslant \alpha_0 \leqslant 6.5 \times 10^{-3}$. Row (a) corresponds to G with $\eta_G = 0.35$ and $\kappa_G = 0.23$, (b) has $\eta_G = 0.35$ and $\kappa_G = 0.33$ and (c) $\eta_G = 0.35$ and $\kappa_G = 0.27$. Even at fixed ecological overlap, different connectances of G give rise to completely different systems in terms of local dynamical stability.

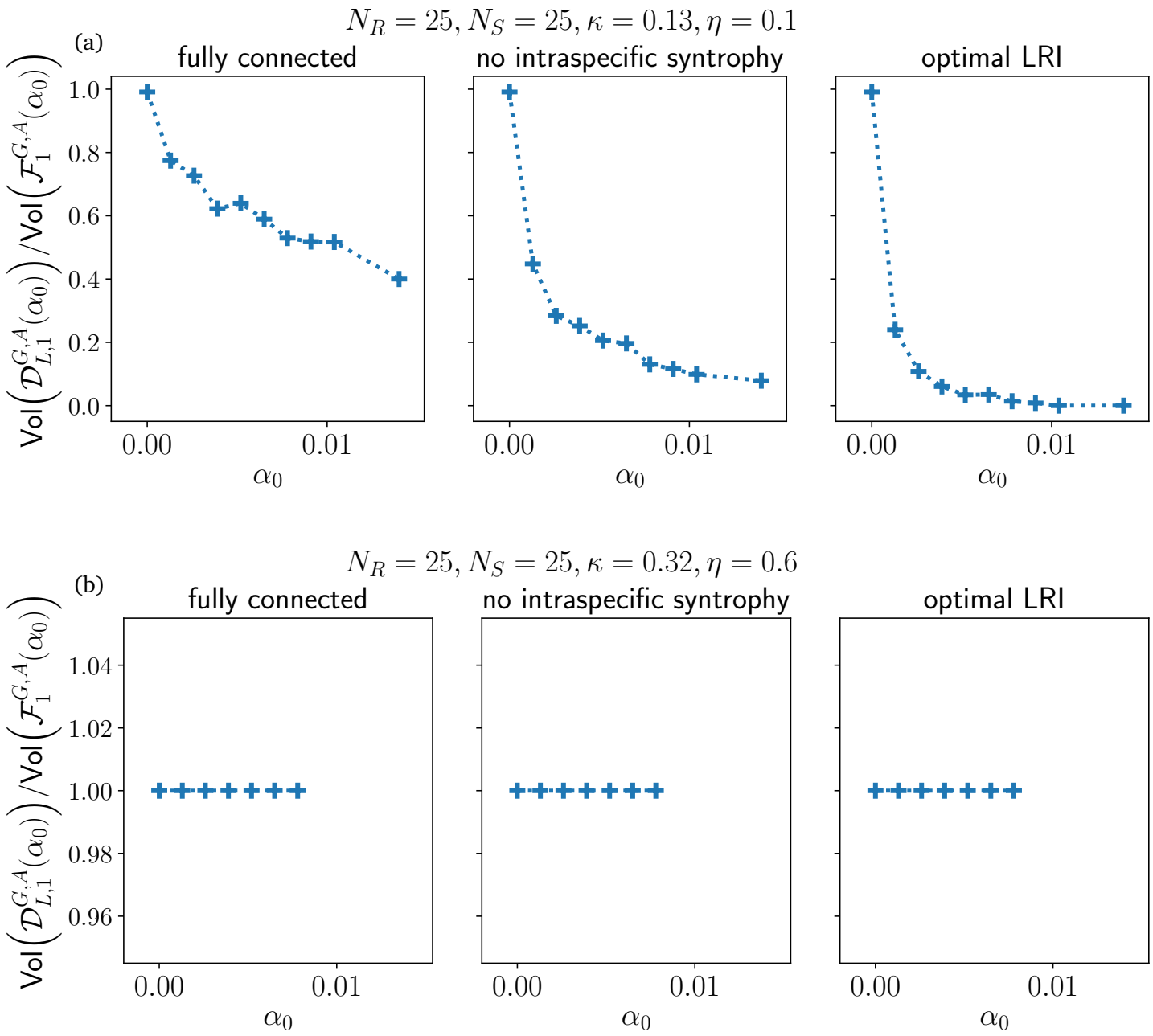

Figure 3.2.6: Ratio of the size of the fully dynamically stable volume and the fully feasible volume for two consumption matrices G (a) with $\eta_G = 0.1$ and $\kappa_G = 0.13$, (b) with $\eta_G = 0.6$ and $\kappa_G = 0.32$. We observe different behaviours for different matrices: for (a) feasibility does not imply local dynamical stability even without syntrophy (it is barely feasible but the ratio is a bit below one for $\alpha_0 = 0$). On the other hand, for (b) feasibility implies local dynamical stability, indeed both regions have the same volume and since $\mathcal{D}_{L,1}^{G,A}(\alpha_0) \subset \mathcal{F}_1^{G,A}(\alpha_0)$, we conclude that both are equal.

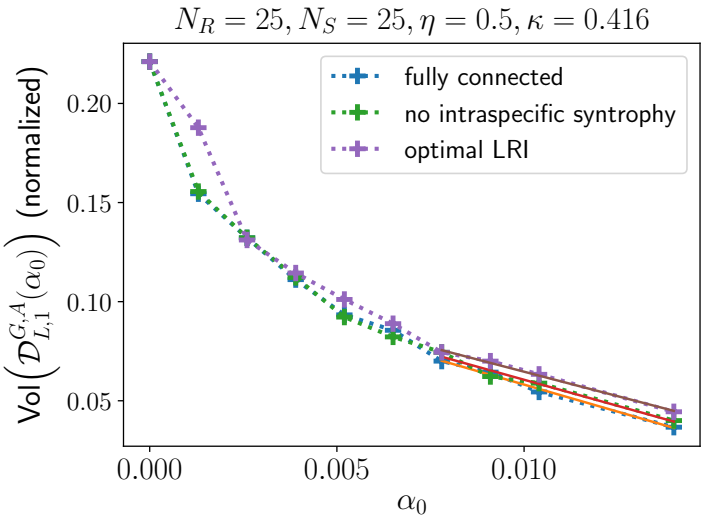

Figure 3.2.7: Evolution of the volume of $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$ with α_0 for a consumption matrix G with $\eta_G = 0.5$ and $\kappa_G = 0.42$. The size of the fully locally dynamically feasible region shrinks as syntrophy increases. A linear fit is performed on the last four points to determine $\alpha_0^D(G, A)$ (see Fig.3.2.8), the point where this curves reaches zero.

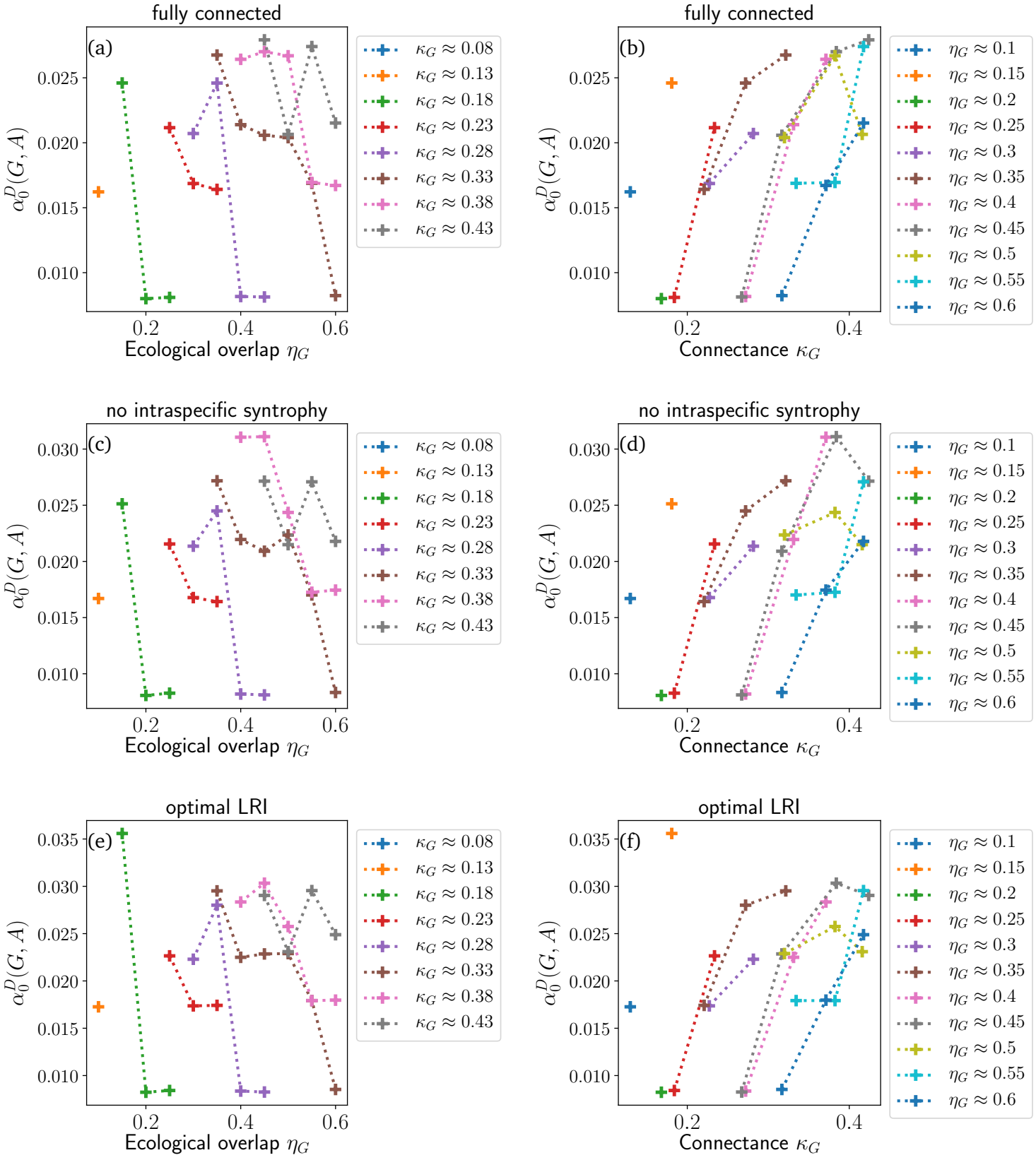

Figure 3.2.8: Critical syntrophy α_0^D, defined as the smallest syntrophy for which we can still find metaparameters that will give rise to fully dynamically stable systems. How α_0^D is estimated is explained in the main text. Errors on α_0^D are not plotted but are at most around 10%. (a)(c)(e) Evolution of α_0^D with ecological overlap η at different connectance. (b)(d)(f) Evolution of α_0^D with connectance κ for different ecological overlap. We observe a strong trend: for a given connectance, α_0^D decreases as ecological overlap increases. Also, for a given ecological overlap, α_0^D increases as connectance is increased.

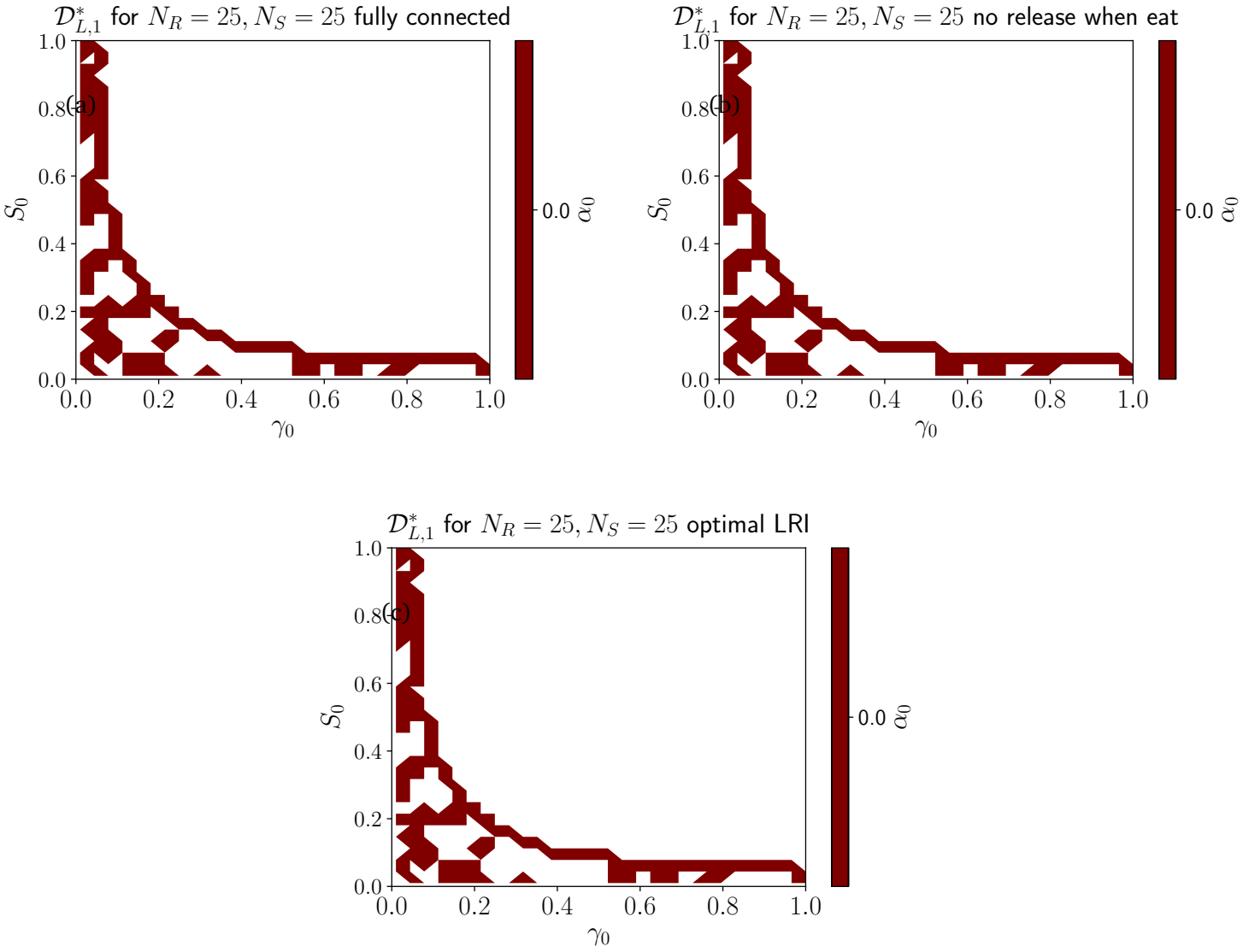

Figure 3.2.9: Common full local dynamical stability volume for different A structures: (a) fully connected, (b) no intraspecific syntrophy and (c) LRI algorithm. The points coloured in dark red give rise to locally dynamically stable systems with probability 1 for all the matrices considered. Very few spots verify this property when there is no syntrophic interaction, and no point gives rise to a fully dynamically stable system for $\alpha_0 = 1.3 \times 10^{-3}$. This is independent of the structure of A that we chose. The white points never give rise to fully dynamically stable systems.

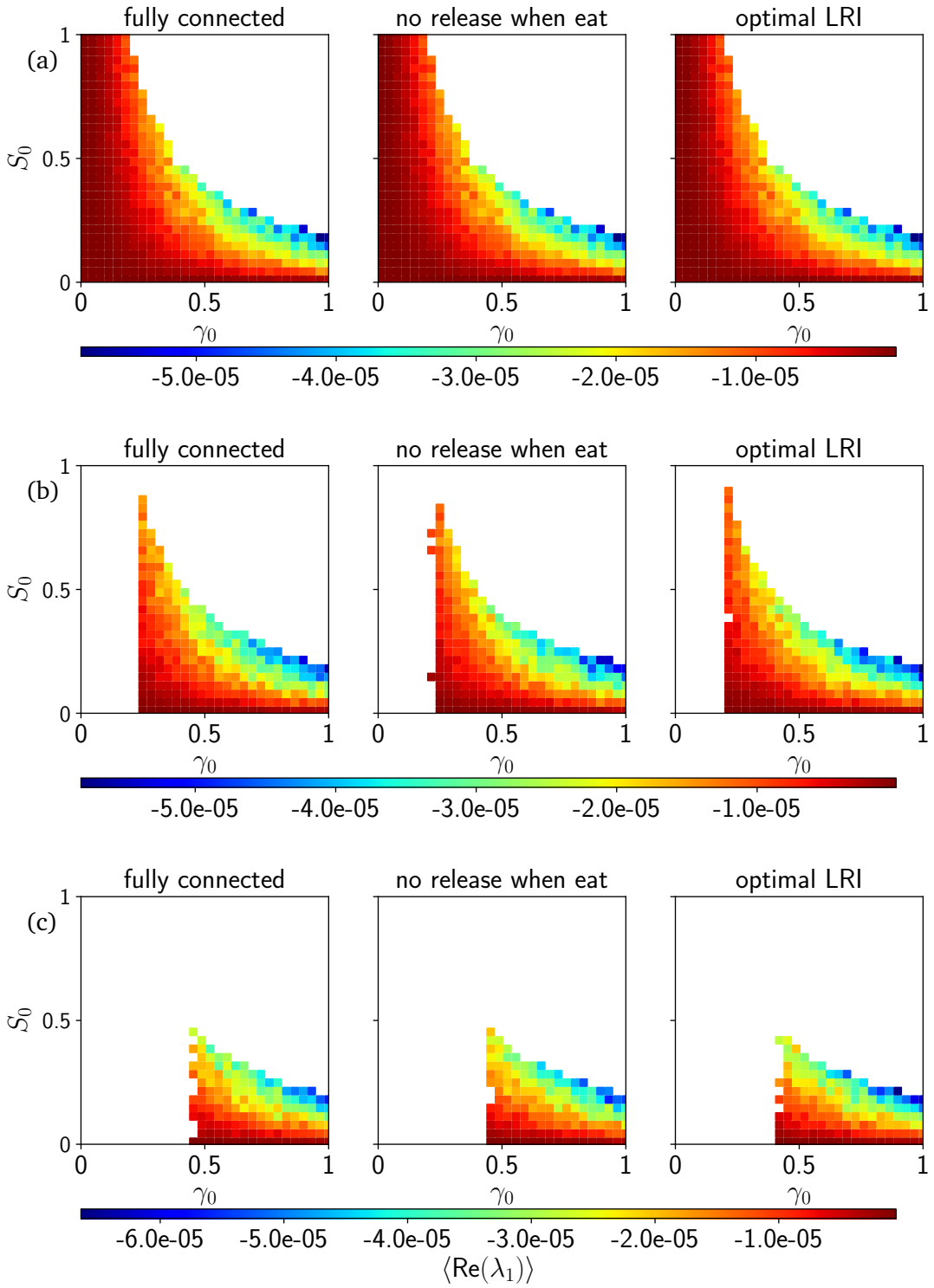

Figure 3.2.10: Largest real eigenvalue $Re(\lambda_1)$ averaged over 300 realisations for each (γ_0, S_0) points for the consumption matrix G with consumers overlap $\eta_G = 0.1$ and connectance $\kappa_G = 0.13$. The white points correspond to not fully dynamically stable systems. Each row corresponds to a different syntrophy value (a) $\alpha_0 = 0$ (no syntrophic interaction), (b) $\alpha_0 = 3.9 \times 10^{-3}$ and (c) $\alpha_0 = 7.8 \times 10^{-3}$. The first column corresponds to the regime where A is fully connected, the second where A forbids intraspecific syntrophy and the third is the outcome of the MC algorithm. As syntrophy increases, the size of the fully dynamically stable region decreases. Furthermore, the boundary points close to the $\gamma_0 \sim S_0^{-1}$ curve are the most stable in every situation.

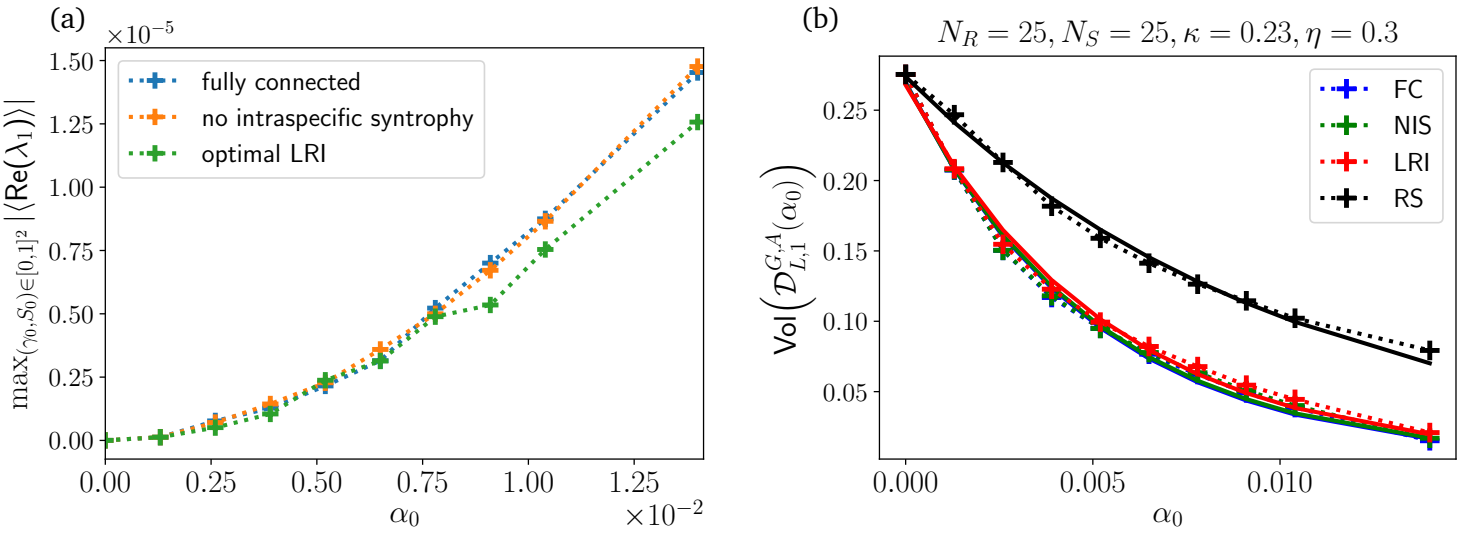

Figure 3.2.11: For a consumption matrix G with $\eta_G = 0.3$ and $\kappa_G = 0.23$. (a) Evolution of the maximal $|\langle Re\,(\lambda_1)\rangle|$ observed in the $(\gamma_0, S_0) \in [0,1]^2$ region. The maximal eigenvalue increases in magnitude, making the system more dynamically stable, as syntrophy increases. That trend is true for all matrices we considered. (b) Volume of $\mathcal{D}^G_{L,1}(\alpha_0)$. As syntrophy increases, fewer and fewer points become fully dynamically stable. For both figures, the different lines show the different stand for the different structure of the syntrophy matrix that we considered.

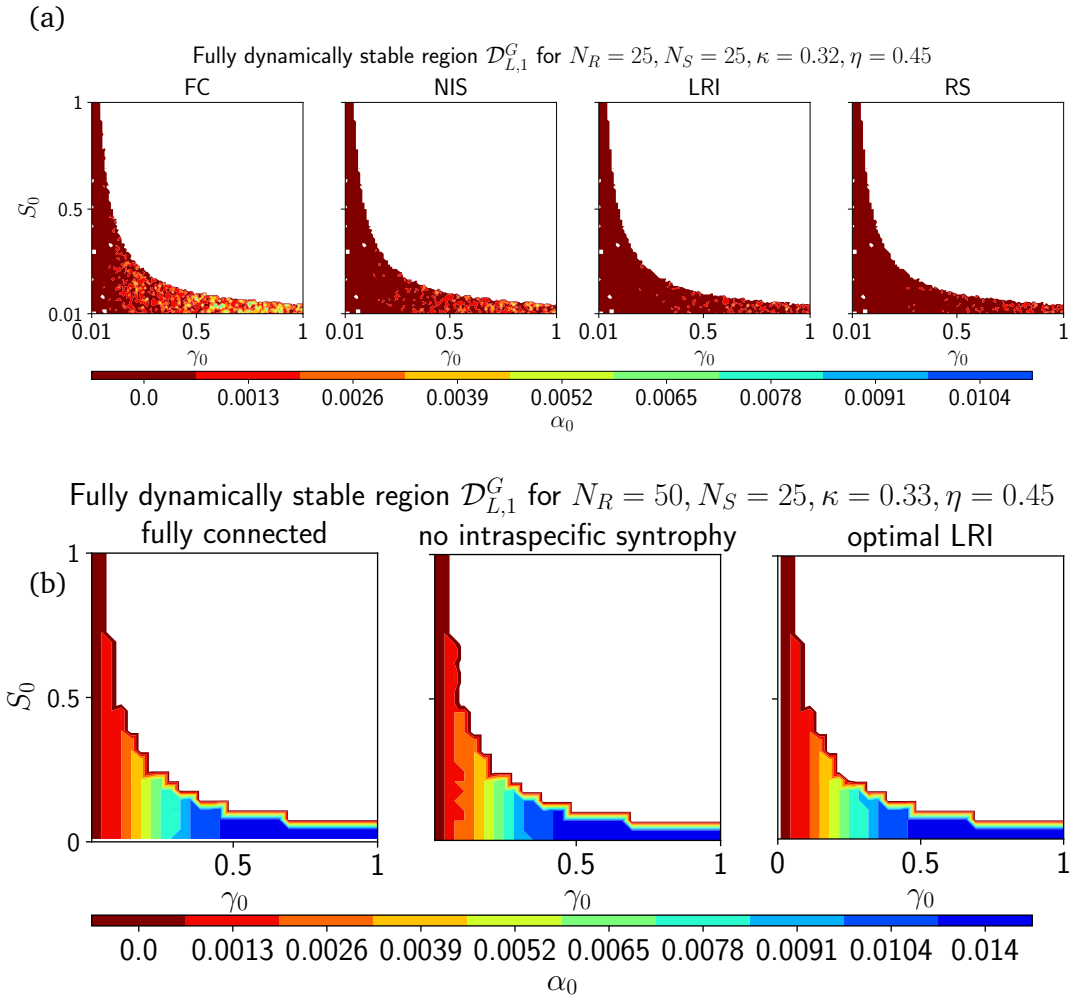

Figure 3.2.12: Fully dynamically stable region $\mathcal{D}^{G,A}_{L,1}$ with the three different structures of A considered: fully connected (left), no intraspecific syntrophy (middle) and LRI matrix (right). The two matrices have the same ecological overlap and connectance, only the number of resources changes. (a) G has $N_R = 25$, $N_S = 25$ and $\kappa_G = 0.32$ and $\eta_G = 0.45$. (b) G has $N_R = 50$, $N_S = 25$ and $\kappa_G = 0.33$ and $\eta_G = 0.45$. The fact that more points can sustain an increased syntrophy is a trend for most of the matrices of the set. ***still check this: the others are more less the same or a tad less.***

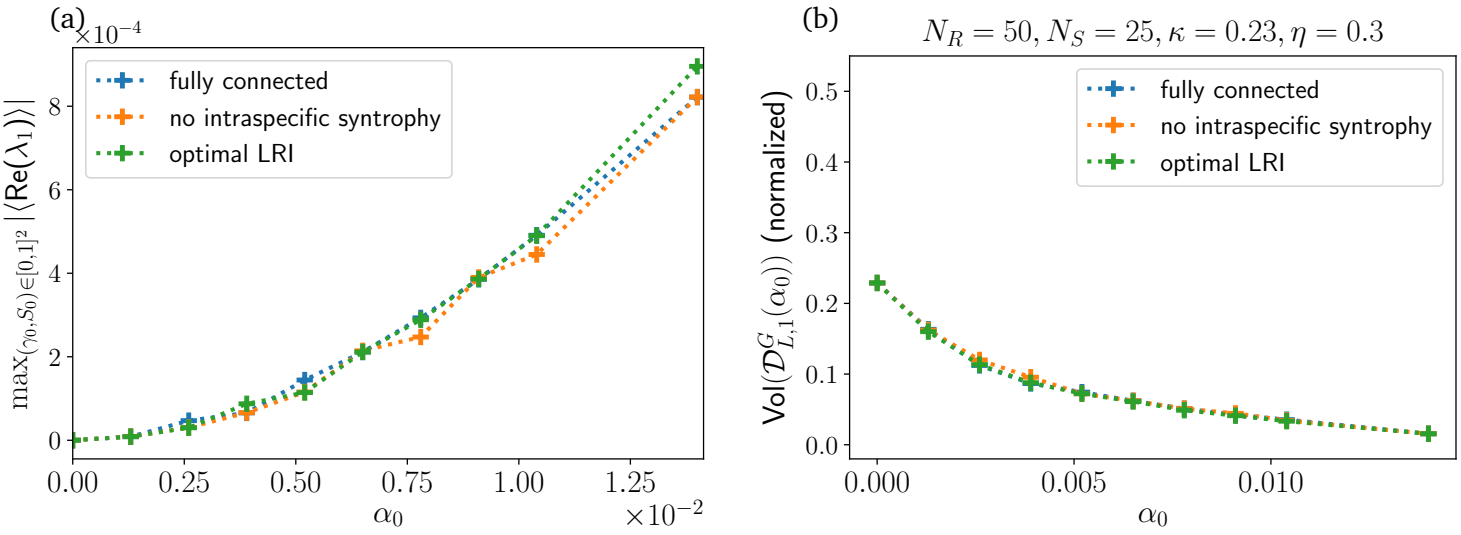

Figure 3.2.13: To be compared with Fig.3.2.11. The consumption matrix G considered here has $\eta_G = 0.3$ and $\kappa_G = 0.23$. (a) Maximal average $|Re\,(\lambda_1)|$ observed in the unit square. (b) Percentage of the unit square occupied by the fully dynamically stable region of G as a function of syntrophy. The matrix considered has almost equal properties to the one in Fig.3.2.11, with the only difference that $N_R = 50$ here. Even though the size of $\mathcal{D}^{G,A}_{L,1}$ is smaller, the eigenvalues are larger in magnitude.

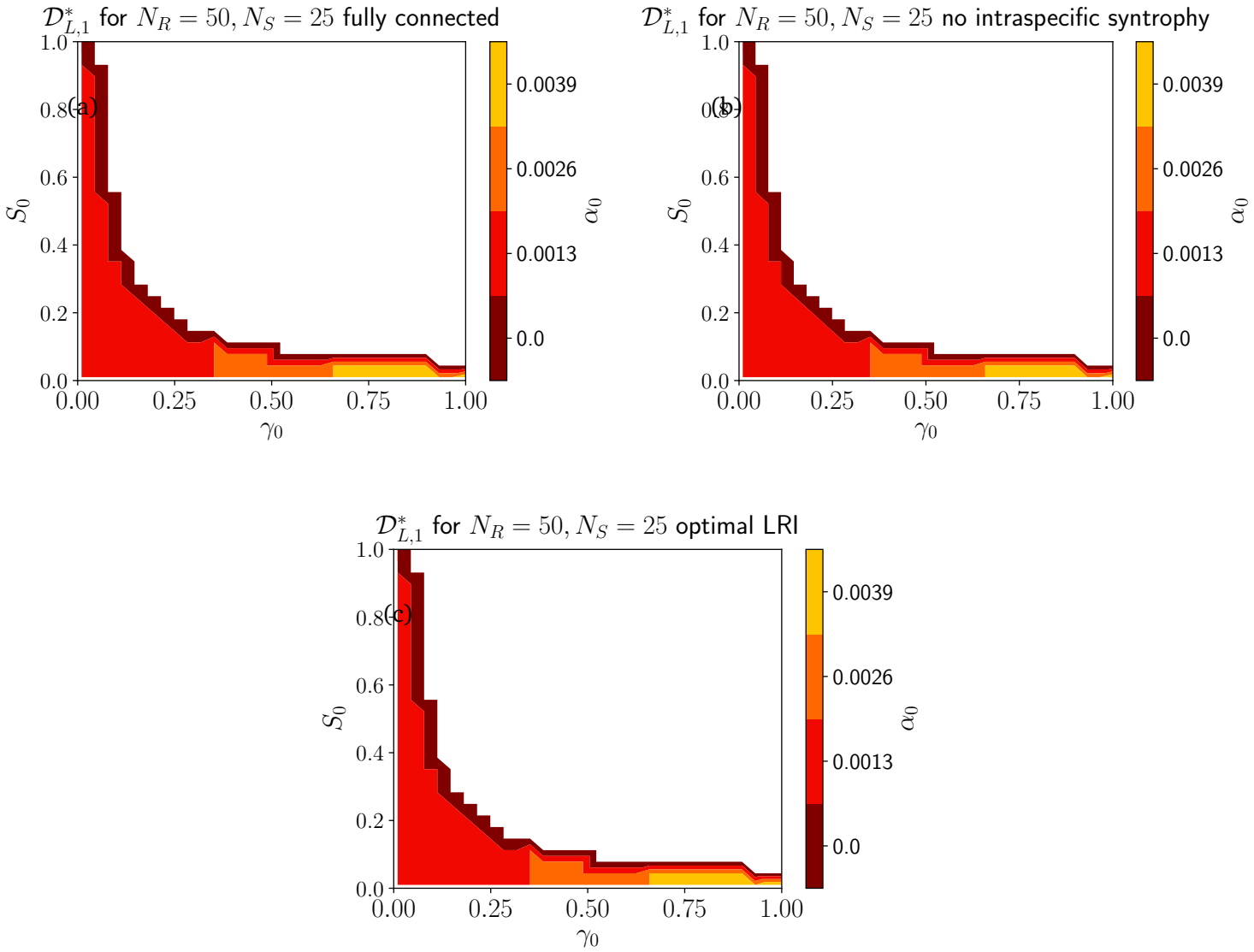

Figure 3.2.14: Common fully dynamically stable volume. It is larger with a larger number of resources -> even though individually it is not always better, it is better for the worse matrices (since the common volume can handle more syntrophy)

3.3 Structural stability

We deal here with the last question of this thesis, which is the one of *structural stability*. We follow the path opened by Bastolla et al. in the Supplementary Material of [22] and later explored by Rohr, Saavedra, and Bascompte in [36]. The latter defined the question of structural stability as *"asking how large is the range of parameter values that are compatible with the stable coexistence of all species"* [36]. Their analysis was done in the context of the study of plants-pollinators networks. For microbial communities, apart from incipient considerations found in [42, 43], only Butler and O'Dwyer fully considered that question [20] and even found a sufficient condition for structural stability of their model under some special assumptions[29].

In the following pages, we investigate how the *critical structural perturbation* $\Delta_S^*(m, G, A)$, which measures for a given set of metaparameters m and a consumption-syntrophy network (G, A) leading to dynamically stable systems how much the external resources feeding rate l_0 may be altered without starting to observe microbial extinctions, changes as a function of the syntrophy α_0 and the structure of both G and A. We aim to find what topologies and syntrophic interactions lead to systems that can sustain the largest perturbations.

3.3.1 Domain of analysis

Appendix 5.2.7 explains the numerical algorithm we designed to determine $\Delta_S^*(m, G, A)$. Although this algorithm always comes to an end, it does not do it fast: typically it takes on the order of ~ 1 hour to provide its result. This means we cannot analyze in detail the whole fully dynamically stable domain $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$ studied earlier and need to focus on some special, most interesting points.

As explained before, in order to assess the structural stability of a given parameter set p, we require that p is locally dynamically stable **say why?**. Hence we need to work with metaparameters m that are in a highly dynamically stable region. Furthermore, in order for comparisons to make sense, $m = (\gamma_0, S_0, \alpha_0)$ should be in a highly dynamically stable region *for every consumption-network*, which is why we choose to study points $m = (\gamma_0, S_0, \alpha_0) \in \mathcal{D}_{L,1}^{G,A}(\alpha_0)$. This gets however a bit tricky: as α_0 increases, the shape of $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$ changes! Since we would like to study how modifying *the syntrophy only* while letting the other parameters fixed impacts the system, we need to choose (γ_0, S_0) such that $(\gamma_0, S_0, \alpha_0) \in \mathcal{D}_{L,1}^{G,A}(\alpha_0)$ for all values of α_0 we want to investigate. The idea is of course to take α_0 as large as possible. Figure 3.2.9 shows that for the case $N_R = N_S = 25$, the largest syntrophy for which $\mathcal{D}_{L,1}^{S_M}(\alpha_0)$ is not empty is smaller than 1.3×10^{-3}, which is only a tenth of the largest common syntrophy ~ 0.01 (Eq.3.2)! Since the case $N_R = N_S = 25$ *a priori* will not provide significant results, we turn to matrices with $N_R > N_S$, which according to literature leads to more stable systems [29]. Indeed Figure 3.2.14 shows that for $N_R = 50$ and $N_S = 25$, $\mathcal{D}_{L,1}^{S_M}(\alpha_0)$ is not empty until at least $\alpha_0 \approx 3.9 \times 10^{-3}$. This indicates that such systems will be more dynamically stable, which is why we choose to work with the set of matrices S_{50} instead of S_{25} **see if name is okay**. Looking at Figure 3.2.14, we see that

[29] The model they consider in [20] differs slightly from ours (*e.g.* resources do not die out $m_\mu = 0$) and the sufficient condition they found assumes fully specialist consumers *i.e.* $N_R = N_S$ and $\gamma = \gamma_0 \mathbb{1}_{N_S}$.

points with $\gamma_0 \gtrapprox 0.7$ can sustain the largest syntrophy while remaining fully dynamically stable for all matrices. In consequence we choose $(\gamma_0, S_0) = (0.75, 0.05)$ and keep these fixed until the end of this section.

Now that we chose γ_0, S_0, N_R and N_S such that we can work with a fairly high syntrophy, we still need to decide for which values of α_0 we compute $\Delta_S^*(m, G, A)$. For a fixed γ_0, S_0, G and A, we define the *critical dynamical syntrophy* $\alpha_C^D(\gamma_0, S_0, G, A)$ as:

$$\alpha_C^D(\gamma_0, S_0, G, A) \equiv \max_{\alpha_0} \{\alpha_0 : \mathcal{D}_{L,1}((\gamma_0, S_0, \alpha_0), G, A) = 1\}. \tag{3.7}$$

In words, $\alpha_C^D(\gamma_0, S_0)$ is the largest syntrophy for which we are sure that a system built with the procedure $\mathcal{A}$ is locally dynamically stable. As is explained in the next section, $\alpha_C^D(\gamma_0, S_0)$ depends heavily on both the structure of G and A and it definitely cannot be approximated as the same for all matrices considered. Since we want to get noticeable effects on $\Delta_S^*(m, G, A)$, we will compute it for each network at its individual critical syntrophy. To add another point of comparison, we will also compute it for each at the lowest critical syntrophy found, which is the largest syntrophy which leads to fully dynamically stable systems for all networks. Finally, we will compare these two Δ_S^* with the one obtained when there is no syntrophy, *i.e.* $\alpha_0 = 0$, which will act as a null model.

3.3.2 Critical dynamical syntrophies

Figure 3.3.1 shows how $\alpha_C^D(\gamma_0, S_0, G, A)$ evolves for the case of A fully connected as a function of connectance κ_G and ecological overlap η_G of the consumption matrix. We observe a clear trend, in accordance with prior results: at fixed ecological overlap, networks with a larger connectance can attain more syntrophy while remaining dynamically stable and at fixed connectance, networks with a larger ecological overlap become unstable faster as syntrophy increases. Apart from the four A structures considered since the beginning of this Thesis (FC, NIS, LRI, RS), we include three additional A-topologies:

- No Intraspecific Syntrophy Consumption matrix Connectance (NISCC); A is random, except that no intraspecific syntrophy is allowed. Its connectance is taken as the connectance of G. This is more or less a "lower connectance version" of the NIS scenario.

- LRI matrix with NIS Connectance (LNISC); A is the outcome of the LRI MCMC procedure described in Methods 2.2.4, except that in contrast to the LRI scenario where $\kappa_A = \kappa_G$, the connectance of A is taken as the one of A in the NIS scenario.

- Random Structure with NIS Connectance (RNISC); A has a completely random structure. Its connectance is chosen as the connectance of A in the NIS regime.

Figure 3.3.2 shows how α_C^D changes as a function of the topology of A. The NISCC and RS outperforms the other scenarios for every consumption matrix considered. Both the FC and NIS cases have κ_A larger than the RS scenario, which hints that systems where many syntrophic interactions take place (*i.e.* A has a large connectance) can sustain an overall smaller maximal syntrophic strength. The main difference between the LRI and the

NISCC/RS regimes, since their respective syntrophy matrix have the same connectance, is their *syntrophic overlap*, *i.e.* the nestedness of A, as is shown by Figure 3.2.3. Although NISCC and RS both have the same connectance and an approximately similar syntrophic overlap, the main difference is that NISCC does not allow intraspecific syntrophy. In the end it seems like dynamical stability is favoured by the following three factors: low connectance of A, low syntrophic overlap of A and prohibition of intraspecific syntrophy. Microbial communities where consumers do not release too many resources – and if they do, in separate niches – can achieve a larger average syntrophy than others while remaining dynamically stable.

3.3.3 Critical structural perturbation

Now that we calculated the critical dynamical syntrophies of each consumption-syntrophy network, we compare their critical structural perturbation $\Delta_S^*(m, G, A)$. As a "null model", we first compute Δ_S^* when there is no syntrophy at play. Figure 3.3.3 shows that structural stability confirms the trend hithertho observed: for a given ecological overlap η_G, Δ_S^* increases as the connectance κ_G increases and for a given κ_G, Δ_S^* decreases as η_G increases. In short, microbial communities where microbes consume a lot of different resources but do not share them resist best to environmental perturbations. These results coincide well with intuition. When a system is structurally perturbed, the external resources feeding rates get shuffled[30], which in turn shuffles the resources available for microbial consumption: some of them start becoming more abundant, some of them get scarcer. If a given microbial species only eats a small number of resources, by luck it is possible that most of its resources got rarer after the environmental perturbation such that the biomass it can eat is not large enough for its survival anymore and it is driven to extinction. On the other hand, if said microbial species eats from many resources, it is unlikely that all the resources it consumes got scarcer after the system perturbation. The lack of biomass from the scarcer resources should indeed be compensated by the additional biomass coming from the more abundant resources, which makes the species less prone to extinction. Having a larger connectance means that on average species consume more resources, which makes the system more stable and at a given connectance, having a larger ecological overlap means that the consumption will have a more triangular shape, which implies there will be some species that eat very few resources which makes the system unstable.

We now focus on what happens when the system is syntrophic, *i.e.* $\alpha_0 > 0$. Figure 3.3.4 shows that surprisingly for the FC, LRI and RS cases, we observe no significant deviation away from the "no syntrophy" case. Whether $\Delta_S^*(m, G, A)$ is computed at each individual α_C^D or at $\min\{\alpha_C^D\}$, it seems like syntrophy does not influence much the structural stability of the system, at least not in a clearly decidable way[31] The only scenario where a clear effect is observable is for Δ_S^* computed at the critical dynamical syntrophy in the NIS regime. There lies an interesting fact: even though an A with a random structure gives the system a larger

[30] By "shuffling", we mean that the l_μ change but their average value does not. Indeed, recall that the ν_μ in Eq.(2.134), have a zero average value which implies that after the perturbation l_0 remains the same, and so should the other metaparameters.

[31] Some effect is sometimes indeed observed but the errors are so large compared to the magnitude of the effect that this very well could be noise.

critical dynamical syntrophy, so in a sense a larger dynamical stability **is this really true**, it does not give it a larger structural stability. We observe a trade-off in the structure of the syntrophy matrix: if you want a larger syntrophy while remaining dynamically stable, you have to give up the fact that it will reinforce your structural stability. On the other hand if you want a configuration such that the syntrophy increases structural stability, the syntrophic strength is not the largest your consumption matrix could theoretically achieve. Benefit gets higher as connectance increases

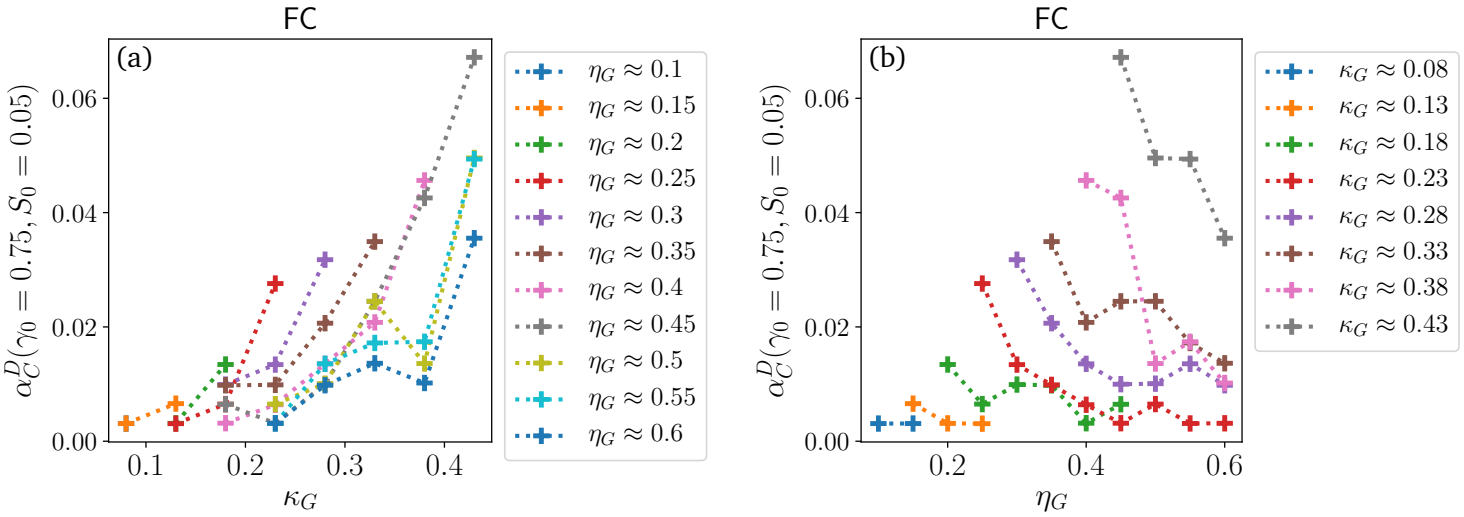

Figure 3.3.1: Critical dynamical syntrophy α_C^D for all the consumption matrices G in the set S_{50} (fully connected syntrophy matrix) (a) as a function of the connectance of G for fixed ecological overlap and (b) as a function of the ecological overlap of G for fixed connectance. For a fixed ecological overlap, systems with a larger connectance can attain larger syntrophies. For a fixed connectance, a small ecological overlap is needed to get a large critical dynamical syntrophy.

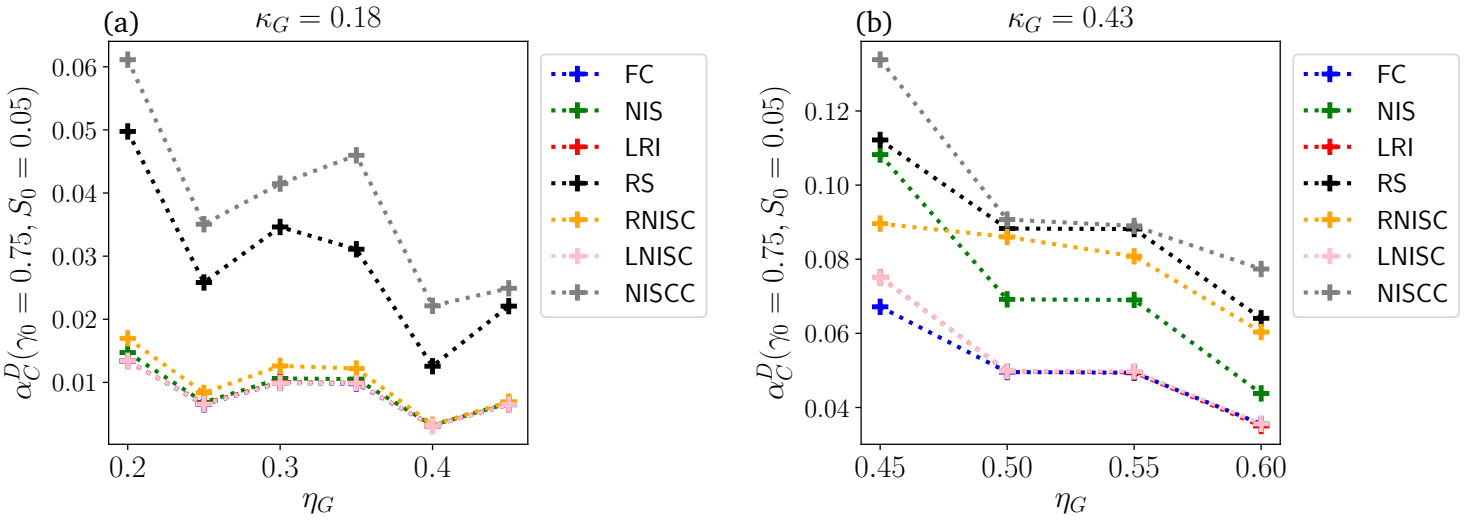

Figure 3.3.2: Critical dynamical syntrophy $\alpha_C^D(\gamma_0 = 0.75, S_0 = 0.05, G, A)$ as a function of the ecological overlap of the consumption matrix G for fixed connectance (a) $\kappa_G = 0.18$ and (b) $\kappa_G = 0.43$. The different lines symbolize the different structures of the syntrophy matrix A: FC, NIS, LRI, RS but also RNISC, LNISC, and NISCC (which are explained in the main text). A higher critical dynamical syntrophy is achieved when A has a random structure. ***because lower connectance??***

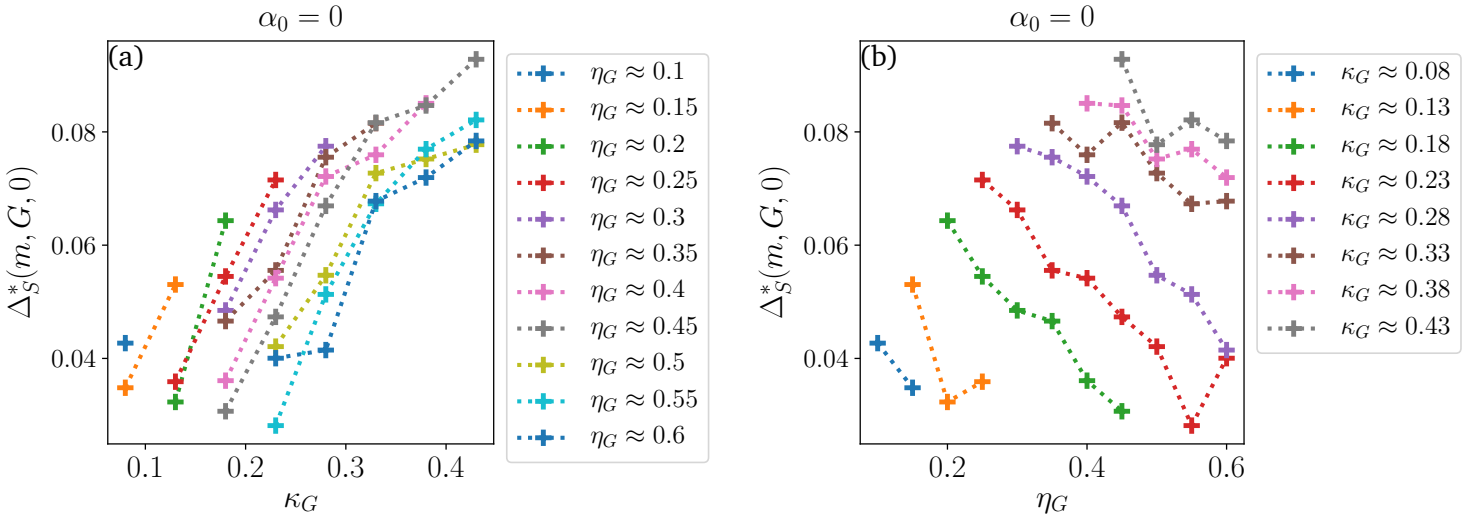

Figure 3.3.3: Critical structural perturbation without syntrophy $\Delta_S^*(m, G, A = 0)$ (a) as a function of ecological overlap with fixed connectance and (b) as a function of connectance for a fixed ecological overlap. We look at matrices with $N_R = 50$ and $N_S = 25$ at one of the points in the metaparameters space that are the most dynamically stable for all the matrices (see Fig.3.2.14), namely $(\gamma_0, S_0) = (0.75, 0.05)$. A clear trend may be observed, which is coherent with what was seen in Figure 3.2.8: for a given connectance, communities with a large ecological overlap are structurally less stable. Similarly, for a given ecological overlap, microbial communities with a consumption matrix with a larger connectance are more structurally stable.

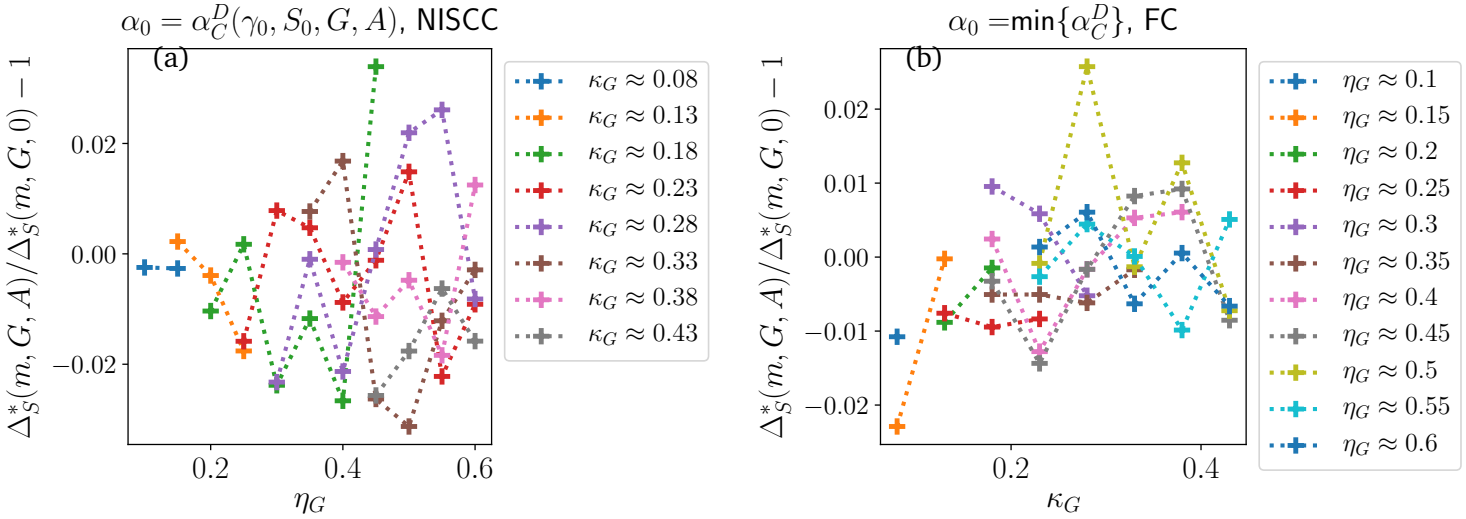

Figure 3.3.4: Typical deviation away from the "no syntrophy" scenario. For all structures of the syntrophy matrix considered, apart from NIS (see Fig.3.3.5), adding a large syntrophy, be it the "common" or "individual" maximum, does not significantly increase the structural stability of the microbial community.

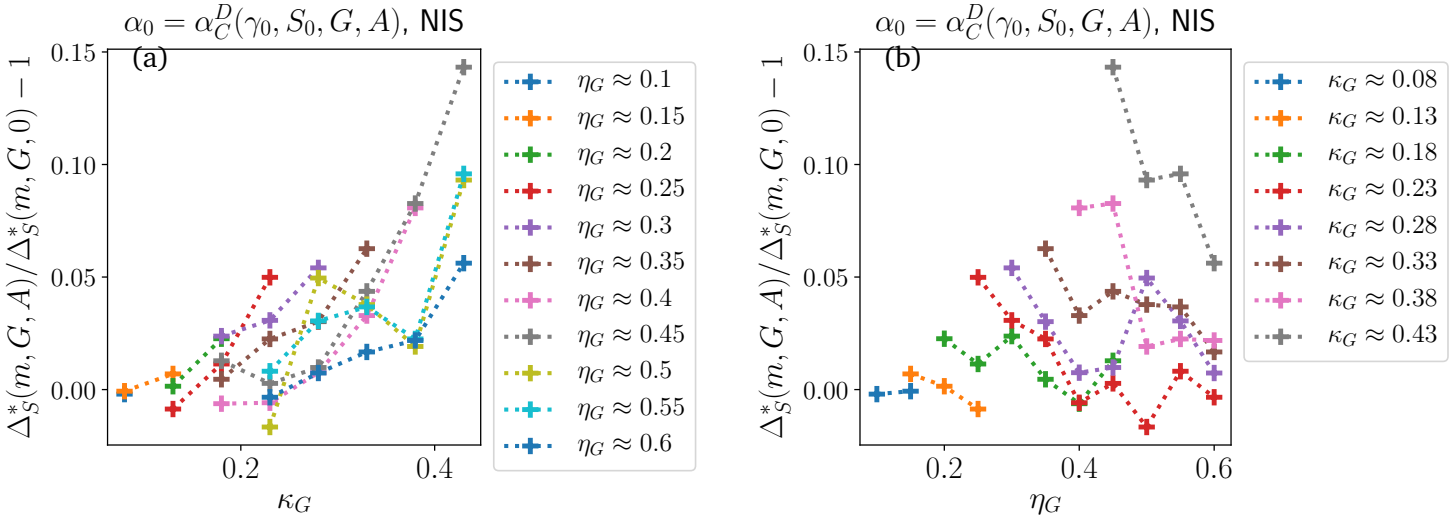

Figure 3.3.5: Deviation from the "no syntrophy" case for the syntrophy matrix A with a NIS structure. α_0 is taken at the critical dynamical syntrophy α_C^D for each system. At the largest possible syntrophy,

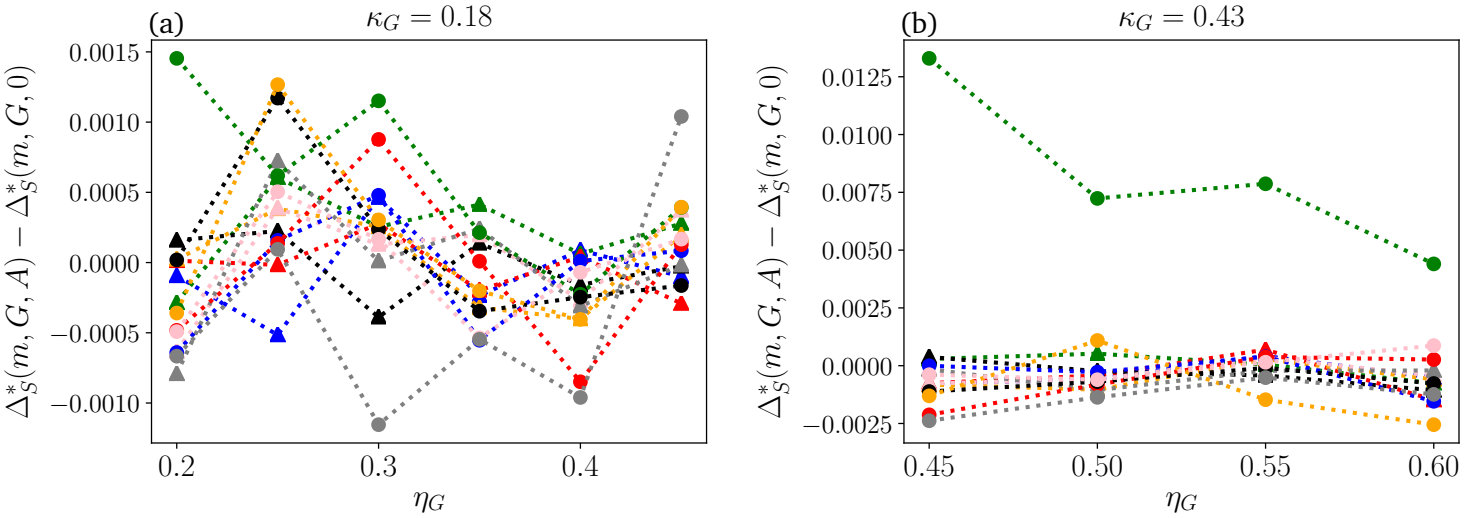

Figure 3.3.6: Difference between the critical structural perturbation with and without syntrophy for different scenarios as a function of the ecological overlap for a given connectance (a) $\kappa_G = 0.18$ and (b) $\kappa_G = 0.43$. The different markers correspond to different α_0 taken : triangle is "common maximum syntrophy", circle is "own individual syntrophy" and square is no syntrophy. The different colours correspond to different structures of A : blue is FC, green NIS, red LRI, black RS, orange RNISC, pink LNISC and grey NISCC. At high connectance, the "own individual syntrophy" NIS scenario is $\sim 10\%$ more structurally stable than the "no syntrophy" case.

4 Discussion

5 Appendices

5.1 Demonstrations

5.1.1 Special determinant computation

We want to know when the determinant of the following N-dimensional square matrix is zero:

$$A_N = \begin{pmatrix} a & b & b \\ b & \ddots & b \\ b & b & a \end{pmatrix}, \textit{i.e. } A_{ij} = b + (a-b)\delta_{ij}. \tag{5.1}$$

The equation we want to solve is:

$$\det\left(A_N\right) = 0. \tag{5.2}$$

Note that, using Gaussian elimination, Eq.(5.2) can be transformed in:

$$\det \begin{pmatrix} a & b & \dots & b \\ b-a & a-b & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & b-a & a-b \end{pmatrix} = 0 \tag{5.3}$$

Using Laplace's expansion, this can be written as:

$$a \det \begin{pmatrix} a-b & 0 & \dots & 0 \\ b-a & a-b & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & b-a & a-b \end{pmatrix} + (a-b) \det \begin{pmatrix} b & b & \dots & b \\ b-a & a-b & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & b-a & a-b \end{pmatrix} = 0 \tag{5.4}$$

Since the first term of the previous equation is a lower triangular matrix, its determinant is easily found:

$$a \det \begin{pmatrix} a-b & 0 & \dots & 0 \\ b-a & a-b & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & b-a & a-b \end{pmatrix} = a\left(a-b\right)^{n-1}. \tag{5.5}$$

Finding an explicit equation for the left term is a bit more involving. Let us define the general n square matrix $F_n(a,b)$:

$$F_n(a,b) = \begin{pmatrix} b & b & \dots & b \\ b-a & a-b & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & b-a & a-b \end{pmatrix}. \tag{5.6}$$

With a Laplace expansion one gets:

$$\det\left(F_n(a,b)\right) = b \det \begin{pmatrix} a-b & 0 & 0 \\ b-a & \ddots & 0 \\ 0 & b-a & a-b \end{pmatrix} + (a-b) \det \begin{pmatrix} b & b & \dots & b \\ b-a & a-b & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & b-a & a-b \end{pmatrix}. \tag{5.7}$$

This means:

$$\det(F_n(a,b)) = b(a-b)^{n-1} + (a-b)\det(F_{n-1}(a,b)). \tag{5.8}$$

It is easy to check that the solution to the previous equation is:

$$\det(F_n(a,b)) = \left[(n-1)b + \det\left(F_1(a,b)\right)\right](a-b)^{n-1}. \tag{5.9}$$

Since $\det\left(F_1(a,b)\right) = 1$, we get:

$$\det\left(F_n(a,b)\right) = n(a-b)^{n-1}b \tag{5.10}$$

Inserting this in Eq.(5.4) yields:

$$\boxed{\det(A_N) = 0 \iff (a-b)^{N-1}\left[a + (N-1)b\right] = 0.} \tag{5.11}$$

5.1.2 The optimal S_0 for locally dynamically stable systems

How S_0 should be adjusted is a bit tricky because it is present in two terms that do not have the same behaviour: l_0^2/S_0 and $N_S^2\alpha_0^2 S_0$. So we need to compute the minimum value the sum of these two terms is and take S_0 as the minimum we found. The consumers equilibrium abundance S_0^* that yields the minimum value is given by

$$\frac{d}{dS_0}\left[\frac{l_0^2}{S_0} + N_S^2\alpha_0^2 S_0\right]_{S_0=S_0^*} = 0 \iff S_0^* = \frac{l_0}{N_S\alpha_0}. \tag{5.12}$$

One checks really easily that this point is indeed a minimum. So if $S_0 > S_0^*$ it should be decreased, and otherwise increased.

5.2 Supplementary material

5.2.1 Why do we solve it this way

TO DO : write this

5.2.2 Algorithmic procedure

We hereby detail the $\mathcal{A}$ procedure which from a set of metaparameters m and a consumption-syntrophy network (G, A) gives rise to a set of parameters p. It goes like this:

1. We first draw randomly each R_ν^* and S_i^* :

$$R_\nu^* = \mathcal{R}\ \forall \nu = 1, \ldots, N_R \text{ and } S_i^* = \mathcal{S}\ \forall i = 1, \ldots, N_S, \tag{5.13}$$

where $\mathcal{R}$ and $\mathcal{S}$ are random variables coming from a distribution of mean equal to the corresponding metaparameter and relative standard deviation[32] ϵ. In our simulations, we chose uniform distributions :

$$\mathcal{R} \sim \text{Unif}(R_0, R_0\epsilon) \text{ and } \mathcal{S} \sim \text{Unif}(S_0, S_0\epsilon). \tag{5.14}$$

[32]By relative standard deviation, we mean the standard deviation measured in units of the average value.

2. The efficiency matrix $\sigma_{i\nu}$ is then drawn similarly, from a distribution with average σ_0. In order to simplify the problem[33], we take a zero-variance like Butler and O'Dwyer in [20], *i.e.* all species consume resources at the same global efficiency :

$$\sigma_{i\nu} = \sigma_0. \tag{5.15}$$

3. We build the consumption matrix $\gamma_{i\nu}$. Its adjacency matrix G is loaded through a user-provided file. While G gives the structure of γ, *i.e.* if a given $\gamma_{i\nu}$ is zero or not, the actual values of $\gamma_{i\nu}$ need then to be determined. They are drawn from a uniform distribution of mean γ_0 and relative standard deviation ϵ:

$$\gamma_{i\nu} = \text{Unif}(\gamma_0, \gamma_0\epsilon)G_{i\nu}. \tag{5.16}$$

4. We draw the resources external feeding rates, similarly to the other parameters :

$$l_\mu = \text{Unif}(l_0, l_0\epsilon)\ \forall\mu = 1, \dots, N_R. \tag{5.17}$$

5. The last free parameters are the non-zero elements of the syntrophy matrix $\alpha_{\nu i}$, since the d_i and l_μ are determined through the equations of evolution at equilibrium. The adjacency matrix A of α needs then to be specified. At the moment, it can be chosen in four different ways : fully connected, in a way that no resource eaten by a given species can be released by that same species (*i.e.* no intraspecific syntrophy : $G_{i\mu} > 0 \iff A_{\mu i} = 0$), random structure with the same number of links as the consumption matrix and finally by a user provided matrix. After the adjacency matrix is loaded, we build α from a uniform distribution of mean α_0 and relative standard deviation ϵ:

$$\alpha_{\nu i} = \text{Unif}(\alpha_0, \alpha_0\epsilon)A_{\nu i}. \tag{5.18}$$

6. With all of these parameters drawn, we can solve Eqs.(2.2a) for the species death rate d_i and Eqs.(2.2b) for m_ν. All the parameters of the model are now fully determined.

7. We check if the constraints Eqs.(2.2) and (2.3) on the parameters are fulfilled. If they are not, we go back to step 1. Otherwise, we can exit the algorithm, a feasible system has been built. We see here the advantage of having input metaparameters that will most likely give rise to a feasible parameter set. If we just give random metaparameters, we run the risk of getting stuck in an unpredictably long loop between steps 1 and 7. If however we are in a region where we know the metaparameters feasibility is high, a feasible system is found much faster.

5.2.3 When is zero part of the spectrum of J^*?

We are interested in knowing when $\lambda = 0$ is part of the spectrum of J^*. By definition, $\lambda = 0$ is an eigenvalue if and only if it solves the master equation (2.47):

$$\det \begin{pmatrix} -\Delta & \Gamma \\ \text{B} & 0 \end{pmatrix} = 0 \tag{5.19}$$

[33]Indeed, we observerd that in general introducing a non uniform σ_0 adds an additional not needed layer of complexity.

Using the fact that Δ is invertible, we can make use of the equality[34]:

$$\det \begin{pmatrix} -\Delta & \Gamma \\ \mathrm{B} & 0 \end{pmatrix} = \det(-\Delta) \det(\mathrm{B}\Delta^{-1}\Gamma). \tag{5.20}$$

Eq.(5.19) then becomes:

$$\det(\mathrm{B}\Delta^{-1}\Gamma) = 0 \tag{5.21}$$

which means that $\mathrm{B}\Delta^{-1}\Gamma$ is not full rank. Finally,

$$\boxed{0 \in \sigma(J^*) \iff \mathrm{B}\Delta^{-1}\Gamma \text{ is not full rank.}} \tag{5.22}$$

But when is $\mathrm{B}\Delta^{-1}\Gamma$ not full rank? Sylvester rank inequality [44] states that:

$$\mathrm{rank}(\mathrm{B}\Delta^{-1}\Gamma) \geqslant \mathrm{rank}(\mathrm{B}) + \mathrm{rank}(\Delta^{-1}\Gamma) - N_R. \tag{5.23}$$

Similarly,

$$\mathrm{rank}(\Delta^{-1}\Gamma) \geqslant \mathrm{rank}(\Delta^{-1}) + \mathrm{rank}(\Gamma) - N_R = \mathrm{rank}(\Gamma), \tag{5.24}$$

where we used the fact that Δ^{-1} is invertible so $\mathrm{rank}(\Delta^{-1}) = N_R$. One of the standard rank properties is:

$$\mathrm{rank}(\Delta^{-1}\Gamma) \leqslant \min\left\{\mathrm{rank}(\Delta^{-1}), \mathrm{rank}(\Gamma)\right\} \implies \mathrm{rank}(\Delta^{-1}\Gamma) \leqslant \mathrm{rank}(\Gamma). \tag{5.25}$$

In the end this yields $\mathrm{rank}(\Delta^{-1}\Gamma) = \mathrm{rank}(\Gamma)$ and :

$$\mathrm{rank}(\mathrm{B}\Delta^{-1}\Gamma) + N_R \geqslant \mathrm{rank}(\mathrm{B}) + \mathrm{rank}(\Gamma). \tag{5.26}$$

We can use this inequality to enunciate a lemma about the presence of zero in the spectrum of J^*.

Lemma 2. *If $N_R \leqslant N_S$ and B and Γ are full rank, then $0 \notin \sigma(J^*)$.*

Proof. We assume that 0 is in the spectrum of J^* and prove it leads to a contradiction. Because $0 \in \sigma(J^*)$, Eq.(5.22) implies $\mathrm{B}\Delta^{-1}\Gamma$ is not full rank. Since the largest possible value for the rank of matrix is the minimum between its number of rows and columns, we have :

$$\mathrm{rank}(\mathrm{B}\Delta^{-1}\Gamma) < \min(N_R, N_S) = N_R. \tag{5.27}$$

This can be used as an upper bound for Eq.(5.26) :

$$2N_R > \mathrm{rank}(\mathrm{B}) + \mathrm{rank}(\Gamma). \tag{5.28}$$

However, we also know that B and Γ are full rank, *i.e.*

$$\mathrm{rank}(\mathrm{B}) = \mathrm{rank}(\Gamma) = N_R. \tag{5.29}$$

Hence the previous inequality amounts to $N_R > N_R$, which is a contradiction. We conclude that the hypothesis $0 \in \sigma(J^*)$ is wrong. □

[34]This uses a formula which is trivially analogous to one found in [31].

5.2.4 Weak LRI regime

Theorem 3. *Let p be a parameter set with a jacobian at equilibrium J^*. If 0 is not an eigenvalue of J^* and the equations*

$$(\Gamma B)_{\mu\mu} < -\sum_{\nu\neq\mu} \left|(\Gamma B)_{\mu\nu}\right| \ \forall\mu, \tag{5.30}$$

are verified, then the real eigenvalues of J^* are negative.

Proof. We assume

$$(\Gamma \mathrm{B})_{\mu\mu} < -\sum_{\nu\neq\mu} \left|(\Gamma \mathrm{B})_{\mu\nu}\right| \ \forall\mu. \tag{5.31}$$

Let $\lambda \in \mathbb{R}$, then the following will also trivially hold:

$$(\Gamma \mathrm{B})_{\mu\mu} - \lambda^2 < -\sum_{\nu\neq\mu} \left|(\Gamma \mathrm{B})_{\mu\nu}\right| \ \forall\mu. \tag{5.32}$$

Dividing Eq.(5.32) by Δ_μ, we get:

$$\frac{1}{\Delta_\mu}\left[\left(\sum_i \Gamma_{\mu i}\mathrm{B}_{i\mu}\right) - \lambda^2\right] < -\sum_{\nu\neq\mu} \left|\frac{\sum_i \Gamma_{\mu i}\mathrm{B}_{i\nu}}{\Delta_\mu}\right| \ \forall\mu. \tag{5.33}$$

Looking at Eq.(2.75), we see that this is equivalent to:

$$S_{\mu\mu} + \sum_{\nu\neq\mu} |S_{\mu\nu}| < 0 \ \forall\mu. \tag{5.34}$$

Using Lemma 1, we know that all the real eigenvalues of $S(\lambda)$ will have a negative real part. We can conclude with the statement of the theorem. □

5.2.5 Effective system

Models which involve the dynamics of species only are in general better known than consumers-resources models [**insert reference**]. In particular, a huge body of literature exists on the study of Lotka-Volterra systems [**insert reference**]. We may profit from this knowledge by transforming the effect of the resources dynamics into an effective consumers-only system.

This can be done by assuming that the resources reach an equilibrium way faster than the consumers. Mathematically, that is equivalent to

$$\frac{dR_\mu}{dt} \approx 0, \forall\mu. \tag{5.35}$$

Using Eq.(1.1a), we get an explicit value for the resources:

$$R_\mu \approx \frac{l_\mu + \sum_j \alpha_{\mu j} S_j}{m_\mu + \sum_k \gamma_{k\mu} S_k}. \tag{5.36}$$

This expression can be used in Eq.(1.1b) to get an effective system which describes the dynamics of the N_S consumers:

$$\frac{dS_i}{dt} = \left(\sum_\nu \left(\frac{\sigma_{i\nu}\gamma_{i\nu}l_\nu}{m_\nu + \sum_k \gamma_{k\nu}S_k} - \alpha_{\nu i} \right) - d_i + \sum_{\nu j} \frac{\sigma_{i\nu}\gamma_{i\nu}\alpha_{\nu j}}{m_\nu + \sum_k \gamma_{k\nu}S_k} S_j \right) S_i. \tag{5.37}$$

This can be rewritten in a more compact way:

$$\frac{dS_i}{dt} = p_i(S)S_i + \sum_j M_{ij}(S)S_iS_j \tag{5.38}$$

with

$$p_i(S) = -\left(d_i + \sum_\nu \alpha_{\nu i} \right) + \sum_\nu \frac{\sigma_{i\nu}\gamma_{i\nu}l_\nu}{m_\nu + \sum_k \gamma_{k\nu}S_k} \text{ and } M_{ij}(S) = \sum_\nu \frac{\sigma_{i\nu}\gamma_{i\nu}\alpha_{\nu j}}{m_\nu + \sum_k \gamma_{k\nu}S_k}. \tag{5.39}$$

If we assume the species S_k are not too far away from their equilibrium values[35], *i.e.*

$$S_k \approx S_k^* \ \forall k, \tag{5.40}$$

then using Eq.(**??**) we can simplify p_i. Indeed,

$$m_\nu + \sum_k \gamma_{k\nu}S_k \approx m_\nu + \sum_k \gamma_{k\nu}S_k^* = \frac{l_\nu + \sum_k \alpha_{\nu k}S_k^*}{R_\nu^*} \tag{5.41}$$

Hence, the explicit dynamical dependence on S can be removed from p_i and M_{ij}:

$$p_i(S) \approx p_i \equiv -\left(d_i + \sum_\nu \alpha_{\nu i} \right) + \sum_\nu \frac{\sigma_{i\nu}\gamma_{i\nu}l_\nu R_\nu^*}{l_\nu + \sum_k \alpha_{\nu k}S_k^*}, \tag{5.42}$$

and

$$M_{ij}(S) \approx M_{ij} \equiv \sum_\nu \frac{\sigma_{i\nu}\gamma_{i\nu}R_\nu^*\alpha_{\nu j}}{l_\nu + \sum_k \alpha_{\nu k}S_k^*}. \tag{5.43}$$

Perturbation analysis

We study a system that we put close to an equilibrium S^*, *i.e.*

$$S = S^* + \Delta S, \text{with } \Delta S \ll 1. \tag{5.44}$$

Written this way, the effective equations of motion Eq.(5.38) are equivalent to:

$$\frac{d\Delta S_i}{dt} = p_i(S^* + \Delta S)\left(S_i^* + \Delta S_i\right) + \sum_j M_{ij}(S^* + \Delta S)\left(S_i^* + \Delta S_i\right)\left(S_j^* + \Delta S_j\right). \tag{5.45}$$

[35]Note that this is very rarely true, especially in the context of the study of structural stability, where entire species sometimes die out.

Since the deviations from equilibrium $\Delta S_i \ll 1$, we can forget the terms in higher power than quadratic:

$$\frac{d\Delta S_i}{dt} = \tilde{p}_i \Delta S_i + \sum_j E_{ij} \Delta S_j + \mathcal{O}(\Delta S^2), \tag{5.46}$$

with

$$\tilde{p}_i \equiv p_i(S^*) + \sum_k M_{ik}(S^*) S_k^*, \tag{5.47}$$

and

$$E_{ij} \equiv \left(\left. \frac{\partial p_i}{\partial S_j} \right|_{S^*} + M_{ij}(S^*) + \sum_k \left. \frac{\partial M_{ik}}{\partial S_j} \right|_{S^*} S_k^* \right) S_i^*. \tag{5.48}$$

After some computations, we can get $\tilde{p}_i$ and E_{ij} in terms of the initial parameters. Indeed,

$$p_i(S^*) = -\left(d_i + \sum_\nu \alpha_{\nu i} \right) + \sum_\nu \frac{\sigma_{i\nu}\gamma_{i\nu} l_\nu}{m_\nu + \sum_k \gamma_{k\nu} S_k^*} \tag{5.49}$$

and

$$M_{ik}(S^*) = \sum_\nu \frac{\sigma_{i\nu}\gamma_{i\nu}\alpha_{\nu j}}{m_\nu + \sum_k \gamma_{k\nu} S_k^*}. \tag{5.50}$$

Hence, using Eq.(5.47):

$$\tilde{p}_i = -\left(d_i + \sum_\nu \alpha_{\nu i} \right) + \sum_\nu \frac{\sigma_{i\nu}\gamma_{i\nu}}{m_\nu + \sum_k \gamma_{k\nu} S_k^*} \left(l_\nu + \sum_j \alpha_{\nu j} S_j^* \right). \tag{5.51}$$

This can be simplified using Eq.(5.41) and Eq.(1.2b):

$$\tilde{p}_i = -d_i + \sum_\nu \sigma_{i\nu}\gamma_{i\nu} R_\nu^* = \sum_\nu \alpha_{\nu i}. \tag{5.52}$$

With a similar computation, one finds

$$E_{ij} = \sum_\nu \frac{\sigma_{i\nu}\gamma_{i\nu} S_i^*}{m_\nu + \sum_k \gamma_{k\nu} S_k^*} \left(\alpha_{\nu j} - \gamma_{j\nu} R_\nu^* \right). \tag{5.53}$$

Finally, Eq.(5.46) can be recast in

$$\frac{d\Delta S_i}{dt} = \sum_j (J_E)_{ij} \Delta S_j, \tag{5.54}$$

where the effective $N_S \times N_S$ jacobian matrix J_E is defined by:

$$(J_E)_{ij} = \sum_\nu \left[\frac{\sigma_{i\nu}\gamma_{i\nu} S_i^*}{m_\nu + \sum_k \gamma_{k\nu} S_k^*} \left(\alpha_{\nu j} - \gamma_{j\nu} R_\nu^* \right) + \alpha_{\nu i}\delta_{ij} \right]. \tag{5.55}$$

We see that we without surprise we find again the B, Γ and Δ matrices coming from the jacobian at equilibrium:

$$(J_E)_{ij} = \sum_\nu \left[\frac{\mathrm{B}_{i\nu}\Gamma_{\nu j}}{\Delta_\nu} + \alpha_{\nu i}\delta_{ij} \right] \tag{5.56}$$

This matrix determines the stability of the equilibrium. Namely if the largest eigenvalue of J_E is positive, the equilibrium is unstable. If it is negative, the equilibrium is stable. If it is zero, the equilibrium is marginal.

5.2.6 Measuring the feasibility and local dynamical stability volumes

In the main text, we study how the sizes of both the fully feasible and the fully locally dynamically stable regions, respectively $\mathcal{F}_1^{G,A}(\alpha_0)$ and $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$, change as a function of G, A and α_0. It is important to explain in what way (and why in *that* way) these are computed. We define the *volume* of $\mathcal{F}_1^{G,A}(\alpha_0)$, or *feasibility volume*, as:

$$\text{Vol}\left(\mathcal{F}_1^{G,A}(\alpha_0)\right) \equiv \frac{\iint_{(\gamma_0,S_0)\in[0,1]^2} d\gamma_0\, dS_0\, \mathcal{F}\left(m,G,A\right)}{\iint_{(\gamma_0,S_0)\in[0,1]^2} d\gamma_0\, dS_0} = \iint\limits_{(\gamma_0,S_0)\in[0,1]^2} d\gamma_0\, dS_0\, \mathcal{F}\left((\gamma_0,S_0,\alpha_0),G,A\right). \tag{5.57}$$

The *volume* of $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$, or *(local) dynamical stability volume* is similarly defined:

$$\text{Vol}\left(\mathcal{D}_{L,1}^{G,A}(\alpha_0)\right) \equiv \iint\limits_{(\gamma_0,S_0)\in[0,1]^2} d\gamma_0\, dS_0\, \mathcal{D}_L\left((\gamma_0,S_0,\alpha_0),G,A\right). \tag{5.58}$$

The feasibility and dynamical stability volumes do not *stricto sensu* measure the area occupied by the fully feasible and fully dynamically stable regions, respectively. In the case of the feasibility volume, it is documented in the main text that the zone where $\mathcal{F}(m,G,A)$ is different from 1 or 0 is very small so measuring the volume this way does not provide a significant difference to naively counting the number of points for which the feasibility function is *exactly* 1.

However, in the case of the dynamical stability volume, this approach would provide unsatisfactory results. Indeed, $\mathcal{D}_L(m,G,A)$ is a function that may be very patchy **TO DO: insert ref to picture of D_L as a function of alpha0**: many points are not fully dynamically stable but rather *almost* fully dynamically stable, in the sense that there are many m such that $\mathcal{D}_L(m,G,A)$ is extremely close to but not exactly equal to 1. Because of this, counting exactly the points where $\mathcal{D}_L(m,G,A)=1$ is very prone to noise, in the sense that it depends a lot on the precision at which $\mathcal{D}_L(m,G,A)$ is evaluated. We know that a lot of points are only *almost* fully dynamically stable; if we increase the number of simulations to evaluate $\mathcal{D}_L(m,G,A)$, who can tell if the points that were previously identified as fully dynamically stable would be now only almost fully dynamically stable and hence not counted in the volume anymore?

Because of such considerations taking into account every point but pondering it by its value of the feasibility/dynamical stability function not only provides a measure close to the idea of measuring the "full volume" but also provides smoother and more robust results.

Following that line of thought, we also define the *common feasibility volume*, which is a measure of the common fully feasible region $\mathcal{F}_1^{S_M}(\alpha_0)$ of a matrix set S_M at a given syntrophic interaction strength α_0:

$$\text{Vol}\left(\mathcal{F}_1^{S_M}\right) \equiv \iint\limits_{(\gamma_0,S_0)\in[0,1]^2} d\gamma_0\, dS_0 \min_{(G,A)\in S_M}\left\{\mathcal{F}\left((\gamma_0,S_0,\alpha_0),G,A\right)\right\}. \tag{5.59}$$

In a completely analogous manner, the *common (local) dynamical stability volume* of a

matrix set S_M is given by:

$$\text{Vol}\left(\mathcal{D}_{L,1}^{S_M}\right) \equiv \iint\limits_{(\gamma_0,S_0)\in[0,1]^2} d\gamma_0\, dS_0 \min_{(G,A)\in S_M} \left\{\mathcal{D}_L\left((\gamma_0,S_0,\alpha_0),G,A\right)\right\}. \tag{5.60}$$

In practice, the integrals appearing in the above formulas are approximated numerically. The unit square $(\gamma_0, S_0) \in [0,1]^2$ is discretized in a set $\{(\gamma_0^i, S_0^i)$ with $i = 1, \ldots, N_{\text{points}}\}$ which are then summed up with their according weights, *e.g.* for Equation (5.57):

$$\text{Vol}\left(\mathcal{F}_1^{G,A}(\alpha_0)\right) \approx \frac{\sum_{i=1}^{N_{\text{points}}} \mathcal{F}\left((\gamma_0^i, S_0^i),\alpha_0,G,A\right)}{N_{\text{points}}}. \tag{5.61}$$

5.2.7 Estimate the critical structural perturbation Δ_S^*

As explained in Methods 2.3.2, $\Delta_S^*(m,G,A)$ solves the equation:

$$P_E\left(\Delta_S^*(m,G,A),m,G,A\right) = 0.5. \tag{5.62}$$

The most straight forward way to find Δ_S^* is to solve Eq.(5.62) numerically. Fortunately, it turns out that, as expected, $P_E(\Delta_S)$ has a typical sigmoidal shape (Fig.5.2.1): the transition between $P_E = 0$ and $P_E = 1$ is sharp compared to the size of the interval $[0,1]$ where Δ_S^* lies. We wrote a solver for Eq.(5.62) that exploits this property. It works in the following way:

1. Through the help of a standard solver from the `GSL` library, find a "high" Δ_H for which $P_E(\Delta_H)$ is very close to 1 but smaller, typically $P_E(\Delta_H) \approx 0.99$. Then find a "low" $\Delta_L < \Delta_H$, very close to 0 but larger.

2. Compute $P_E(\Delta_S)$ for N_{points} points Δ_S homogeneously spread in the interval $[\Delta_L, \Delta_H]$.

3. Because the $P_E(\Delta_S)$ computed at the previous step typically form a sigmoidal shape, fit these points with a sigmoidal function $S_f(\Delta_S)$. We choose:

$$S_f(\Delta_S) \equiv \frac{1}{1+e^{-C_1(\Delta_S - C_2)}}, \tag{5.63}$$

   where the constants C_1, C_2 precisely are estimated through the fitting procedure.

4. $\Delta_S^*(m,G,A)$ is obtained by solving analytically $S_f(\Delta_S^*) = 0.5$. For the choice of Eq.(5.63), this is trivial : $\Delta_S^* = C_2$. Indeed $S_f(C_2) = 1/(1+1) = 0.5$. We take the error on C_2 obtained through the standard fitting routine from the `GSL` library as the "error" on Δ_S^*.

Finally, it is worth mentioning that $P_E(\Delta_S, m, G, A)$, which again is the probability to observe *at least* one extinction when structurally perturbing a parameters set $\mathcal{A}(m,G,A)$ by Δ_S, is estimated numerically through the following procedure:

1. Create a parameters set $\mathcal{A}(m,G,A)$.

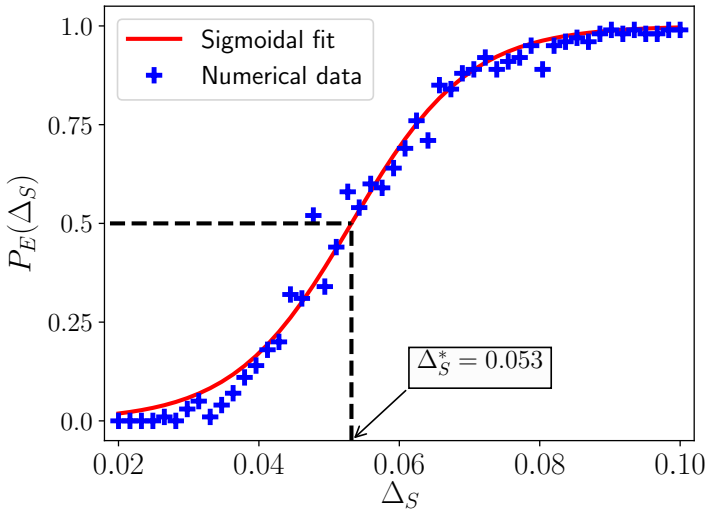

Figure 5.2.1: Typical probability of finding one extinction when structurally perturbing the system with a magnitude Δ_S. The critical structural perturbation is easily estimated with a sigmoidal fit.

2. Structurally perturb it by an amount Δ_S.
3. Time evolve the parameters set until an equilibrium is reached. Compute $p_E \in \{0; 1\}$, which is 0 if no extinction has been observed, and 1 if at least one extinction occurred.
4. Repeat steps 1–3 N_{sys} times. $P_E(\Delta_S)$ is the average value of p_E.

For the figures of Results 3.3, we used $N_{\text{points}} = 125$ and $N_{\text{sys}} = 50$. We observed (although did not have the time to quantify it properly) that increasing N_{points} reduces the error on Δ_S^* faster than increasing N_{sys}.

Bibliography

[1] Ashutosh Jogalekar. *Physicists in Biology; And Other Quirks of the Genomic Age*. en. Library Catalog: blogs.scientificamerican.com.

[2] Erwin Schrödinger. *What is Life? The Physical Aspect of the Living Cell*. Cambridge University Press, 1944.

[3] Hue Sun Chan and Ken A Dill. "The Protein Folding Problem". en. In: (1993), p. 10.

[4] C. K. Fisher and P. Mehta. "The transition between the niche and neutral regimes in ecology". en. In: *Proceedings of the National Academy of Sciences* 111.36 (Sept. 2014), pp. 13111–13116. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1405637111.

[5] Alessandro Attanasi et al. "Information transfer and behavioural inertia in starling flocks". en. In: *Nature Physics* 10.9 (Sept. 2014), pp. 691–696. ISSN: 1745-2473, 1745-2481. DOI: 10.1038/nphys3035.

[6] Ruth E. Ley et al. "Worlds within worlds: evolution of the vertebrate gut microbiota". en. In: *Nature Reviews Microbiology* 6.10 (Oct. 2008), pp. 776–788. ISSN: 1740-1526, 1740-1534. DOI: 10.1038/nrmicro1978.

[7] Cristina Becerra-Castro et al. "Wastewater reuse in irrigation: A microbiological perspective on implications in soil fertility and human and environmental health". en. In: *Environment International* 75 (Feb. 2015), pp. 117–135. ISSN: 01604120. DOI: 10.1016/j.envint.2014.11.001.

[8] P. G. Falkowski. "Biogeochemical Controls and Feedbacks on Ocean Primary Production". en. In: *Science* 281.5374 (July 1998), pp. 200–206. DOI: 10.1126/science.281.5374.200.

[9] A J Lotka. "Analytical Note on Certain Rhythmic Relations in Organic Systems". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 6.7 (July 1920), pp. 410–415. ISSN: 0027-8424. DOI: 10.1073/pnas.6.7.410.

[10] Babak Momeni, Li Xie, and Wenying Shou. "Lotka-Volterra pairwise modeling fails to capture diverse pairwise microbial interactions". en. In: *eLife* 6 (Mar. 2017), e25051. ISSN: 2050-084X. DOI: 10.7554/eLife.25051.

[11] James P. O'Dwyer. "Whence Lotka-Volterra?: Conservation laws and integrable systems in ecology". en. In: *Theoretical Ecology* 11.4 (Dec. 2018), pp. 441–452. ISSN: 1874-1738, 1874-1746. DOI: 10.1007/s12080-018-0377-0.

[12] Robert MacArthur. "Species packing and competitive equilibrium for many species". en. In: *Theoretical Population Biology* 1.1 (May 1970), pp. 1–11. ISSN: 00405809. DOI: 10.1016/0040-5809(70)90039-0.

[13] James D. Brunner and Nicholas Chia. "Metabolite mediated modeling of microbial community dynamics captures emergent behavior more effectively than species-species modeling". en. In: *Journal of The Royal Society Interface* 16.159 (Oct. 2019). arXiv: 1907.04436, p. 20190423. ISSN: 1742-5689, 1742-5662. DOI: 10.1098/rsif.2019.0423.

[14] Brandon E.L. Morris et al. "Microbial syntrophy: interaction for the common good". en. In: *FEMS Microbiology Reviews* 37.3 (May 2013), pp. 384–406. ISSN: 1574-6976. DOI: 10.1111/1574-6976.12019.

[15] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. “What is flux balance analysis?” en. In: *Nature Biotechnology* 28.3 (Mar. 2010), pp. 245–248. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/nbt.1614.

[16] Ines Thiele et al. “Multiscale Modeling of Metabolism and Macromolecular Synthesis in E. coli and Its Application to the Evolution of Codon Usage”. en. In: *PLoS ONE* 7.9 (Sept. 2012). Ed. by Tamir Tuller, e45635. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0045635.

[17] Pietro Landi et al. “Complexity and stability of ecological networks: a review of the theory”. en. In: *Population Ecology* 60.4 (Oct. 2018), pp. 319–345. ISSN: 1438-3896, 1438-390X. DOI: 10.1007/s10144-018-0628-3.

[18] T W James. “Continuous Culture of Microorganisms”. en. In: *Annual Review of Microbiology* 15.1 (Oct. 1961), pp. 27–46. ISSN: 0066-4227, 1545-3251. DOI: 10.1146/annurev.mi.15.100161.000331.

[19] Alberto Pascual-García and Ugo Bastolla. “Mutualism supports biodiversity when the direct competition is weak”. en. In: *Nature Communications* 8.1 (Apr. 2017), p. 14326. ISSN: 2041-1723. DOI: 10.1038/ncomms14326.

[20] Stacey Butler and James P. O’Dwyer. “Stability criteria for complex microbial communities”. en. In: *Nature Communications* 9.1 (Dec. 2018), p. 2970. ISSN: 2041-1723. DOI: 10.1038/s41467-018-05308-z.

[21] Mimmo Iannelli and Andrea Pugliese. *An Introduction to Mathematical Population Dynamics*. en. Vol. 79. UNITEXT. Cham: Springer International Publishing, 2014. ISBN: 978-3-319-03025-8 978-3-319-03026-5. DOI: 10.1007/978-3-319-03026-5.

[22] Ugo Bastolla et al. “The architecture of mutualistic networks minimizes competition and increases biodiversity”. en. In: *Nature* 458.7241 (Apr. 2009), pp. 1018–1020. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature07950.

[23] Samuel Jonhson, Virginia Domínguez-García, and Miguel A. Muñoz. “Factors Determining Nestedness in Complex Networks”. en. In: *PLoS ONE* 8.9 (Sept. 2013). Ed. by Yamir Moreno, e74025. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0074025.

[24] Robert M. May. “Will a Large Complex System be Stable?” In: *Nature* 238.5364 (Aug. 1972), pp. 413–414. ISSN: 1476-4687. DOI: 10.1038/238413a0.

[25] Stefano Allesina and Si Tang. “Stability Criteria for Complex Ecosystems”. en. In: *Nature* 483.7388 (Mar. 2012). arXiv: 1105.2071, pp. 205–208. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature10832.

[26] Stefano Allesina et al. “Predicting the stability of large structured food webs”. en. In: *Nature Communications* 6.1 (Nov. 2015), p. 7842. ISSN: 2041-1723. DOI: 10.1038/ncomms8842.

[27] Matthieu Barbier and Jean-François Arnoldi. *The cavity method for community ecology*. en. preprint. Ecology, June 2017. DOI: 10.1101/147728.

[28] A. R. Ives and S. R. Carpenter. “Stability and Diversity of Ecosystems”. en. In: *Science* 317.5834 (July 2007), pp. 58–62. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1133258.

[29] Giulio Biroli, Guy Bunin, and Chiara Cammarota. “Marginally stable equilibria in critical ecosystems”. en. In: *New Journal of Physics* 20.8 (Aug. 2018), p. 083051. ISSN: 1367-2630. DOI: 10.1088/1367-2630/aada58.

[30] Oscar Perron. “Zur Theorie der Matrices”. de. In: (), p. 16.

[31] Philip D. Powell. “Calculating Determinants of Block Matrices”. en. In: *arXiv:1112.4379 [math]* (Dec. 2011). arXiv: 1112.4379.

[32] Sergej Gerschgorin. “"Uber die Abgrenzung der Eigenwerte einer Matrix”. de. In: *Bulletin de l’Académie des Sciences de l’URSS. Classe des sciences mathématiques et na* 6 (1931), pp. 749–754.

[33] Daniel Delahaye, Supatcha Chaimatanan, and Marcel Mongeau. “Simulated Annealing: From Basics to Applications”. en. In: *Handbook of Metaheuristics*. Ed. by Michel Gendreau and Jean-Yves Potvin. Vol. 272. Series Title: International Series in Operations Research & Management Science. Cham: Springer International Publishing, 2019, pp. 1–35. ISBN: 978-3-319-91085-7 978-3-319-91086-4. DOI: 10.1007/978-3-319-91086-4_1.

[34] Jan R. Magnus and Heinz Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. en. Third edition. Wiley series in probability and statistics. Hoboken, NJ: Wiley, 2019. ISBN: 978-1-119-54119-6 978-1-119-54116-5.

[35] Stefano Allesina and Si Tang. “The stability–complexity relationship at age 40: a random matrix perspective”. en. In: *Population Ecology* 57.1 (Jan. 2015). Publisher: John Wiley & Sons, Ltd, pp. 63–75. ISSN: 1438-390X. DOI: 10.1007/s10144-014-0471-0.

[36] R. P. Rohr, S. Saavedra, and J. Bascompte. “On the structural stability of mutualistic systems”. en. In: *Science* 345.6195 (July 2014), pp. 1253497–1253497. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1253497.

[37] J. Bascompte et al. “The nested assembly of plant-animal mutualistic networks”. en. In: *Proceedings of the National Academy of Sciences* 100.16 (Aug. 2003), pp. 9383–9387. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1633576100.

[38] M. B. Bonsall. “Life History Trade-Offs Assemble Ecological Guilds”. en. In: *Science* 306.5693 (Oct. 2004), pp. 111–114. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1100680.

[39] E. Thebault and C. Fontaine. “Stability of Ecological Communities and the Architecture of Mutualistic and Trophic Networks”. en. In: *Science* 329.5993 (Aug. 2010), pp. 853–856. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1188321.

[40] G. Hardin. “The Competitive Exclusion Principle”. en. In: *Science* 131.3409 (Apr. 1960), pp. 1292–1297. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.131.3409.1292.

[41] Xin Wang and Yang-Yu Liu. “Overcome Competitive Exclusion in Ecosystems”. en. In: *arXiv:1805.06002 [physics, q-bio]* (May 2019). arXiv: 1805.06002.

[42] Mikhail Tikhonov and Remi Monasson. “Collective Phase in Resource Competition in a Highly Diverse Ecosystem”. en. In: *Physical Review Letters* 118.4 (Jan. 2017), p. 048103. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.118.048103.

[43] Robert Marsland et al. “Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities”. en. In: *PLOS Computational Biology* 15.2 (Feb. 2019). Ed. by Alexandre V. Morozov, e1006793. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006793.

[44] Néstor Thome. “Inequalities and equalities for $l = 2$ (Sylvester), $l = 3$ (Frobenius), and $l > 3$ matrices”. en. In: *Aequationes mathematicae* 90.5 (Oct. 2016), pp. 951–960. ISSN: 0001-9054, 1420-8903. DOI: 10.1007/s00010-016-0412-4.