

# Syntrophic interaction in microbial communities

Léo Buchenel

May 6, 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

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Consumer Resource Models in microbial ecology	5
1.2	Establishing the model and goals	5
1.2.1	Equilibria of the model	7
1.2.2	Attack strategy and important notions	7
<b>2</b>	<b>Methods</b>	<b>11</b>
2.1	Feasibility	11
2.1.1	Basic concepts	11
2.1.2	The feasibility region	13
2.1.3	Estimating the fully feasible region $\mathcal{F}_1^{G,A}$	14
2.2	Dynamical stability	17
2.2.1	Definitions	17
2.2.2	The quest for a full solution	20
2.2.3	Bounds on the eigenvalues	21
2.2.4	Low intra resources interaction (LRI) regime	24
2.2.5	Application: fully connected consumption and syntrophy networks	29
2.3	Structural stability	34
2.3.1	Definitions	34
2.3.2	Numerical estimate of the critical structural perturbation	35
<b>3</b>	<b>Results</b>	<b>37</b>
3.1	Feasibility	38
3.1.1	The feasibility region in the absence of syntrophy	38
3.1.2	Impact of syntrophy on the feasible region	40
3.2	Dynamical stability	52
3.2.1	LRI regime – Outcome of the MCMC algorithm	52
3.2.2	Typical dynamical stability	53
3.2.3	Fully dynamically stable region	53
3.2.4	Largest eigenvalue of the jacobian	56
3.2.5	The influence of the matrix dimension	57
3.3	Structural stability	73
3.3.1	Domain of analysis	73
3.3.2	Critical dynamical syntrophies	74
3.3.3	Critical structural perturbation	75
<b>4</b>	<b>Discussion</b>	<b>81</b>
<b>5</b>	<b>Appendices</b>	<b>85</b>
5.1	Demonstrations	85
5.1.1	Special determinant computation	85
5.1.2	The optimal $S_0$ for locally dynamically stable systems	86
5.2	Supplementary material	86
5.2.1	Why do we solve it this way	86

5.2.2	Algorithmic procedure	87
5.2.3	When is zero part of the spectrum of $J^*$ ?	88
5.2.4	Weak LRI regime	89
5.2.5	Effective system	90
5.2.6	Measuring the feasibility and local dynamical stability volumes	92
5.2.7	Center of dynamical stability	93
5.2.8	Probability of dynamical stability when feasibility is ensured	94
5.2.9	Estimate the critical structural perturbation $\Delta_S^*$	95
	<b>Bibliography</b>	<b>97</b>

# 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 contemporary 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 vertebrates' 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<sup>1</sup> 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<sup>2</sup>. Resources are labelled  $\mu = 1, \dots, N_R$

---

<sup>1</sup>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.

<sup>2</sup>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:

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

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

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 referred 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  $\mu$  has abundance  $R_\mu$ ). 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  $\mu$  has death rate  $m_\mu$ ).
- 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  $\mu$ ).

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  $R_\mu$  of a resource  $\mu$  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  $m_\mu$ : 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  $-\sum_j \gamma_{j\mu} R_\mu S_j$ ,
- Intrasytemic 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  $R_\nu$  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  $\nu$  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*<sup>3</sup> of our model if they are fixed points of it, that means if the following equations are fulfilled:

$$\begin{cases} 0 = l_\mu - m_\mu R_\mu^* - \sum_j \gamma_{j\mu} R_\mu^* S_j^* + \sum_j \alpha_{\mu j} S_j^* \end{cases} \quad (1.2a)$$

$$\begin{cases} 0 = \sum_\nu \sigma_{i\nu} \gamma_{i\nu} R_\nu^* S_i^* - d_i S_i^* - \sum_\nu \alpha_{\nu i} S_i^* \end{cases} \quad (1.2b)$$

### 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 complementary 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:

$$\mathcal{P} \equiv \{p : p = (l_\mu, m_\mu, d_i, \gamma_{i\mu}, \alpha_{\mu i}, \sigma_{i\mu}, R_\mu^*, S_i^*)\} \quad (1.3)$$

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]:

$$q_{i\mu} = \Omega Q_{i\mu} \quad (1.4)$$

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

<sup>3</sup>For the sake of brevity, we will sometimes drop the  $\mu$  and  $j$  subscripts and simply write  $\{R^*, S^*\}$ .

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

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

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  $Q$  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:

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

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*:

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

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*  $\mathcal{C} : \mathcal{P} \rightarrow \mathcal{M} \times \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S}$  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  $3N_R + 2N_S + 3N_R N_S$  (continuous) to 8 (continuous) +  $3N_R N_S$  (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**<sup>4</sup>, this measures how “nested” the system is, *i.e.* if there are clusters grouped together<sup>5</sup>. It is known [19, 22] that nestedness can play a profound role in the dynamics of ecological communities. Although it is somewhat controversial [23], we will keep the definition of the nestedness  $\eta(M)$  of a binary matrix  $M$  as it was used in [22]:

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

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

$$n_i \equiv \sum_k M_{ik}, \quad (1.9)$$

and  $n_{ij}$  is the overlap matrix defined as

$$n_{ij} \equiv \sum_k M_{ik} M_{jk}. \quad (1.10)$$

- 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:

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

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

$$\mathcal{A} : \mathcal{M} \times \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S} \rightarrow \mathcal{P} \quad (1.12)$$

which brings us from the collapsed space to the parameter space<sup>6</sup>. 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.

<sup>4</sup>For the matrix consumption  $G$ , we will call it especially the “ecological overlap”.

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

<sup>6</sup>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  $\alpha_{\mu i} \geq 0$  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<sup>7</sup>: 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:

$$\boxed{R_\mu^*, S_i^*, \sigma_{i\mu}, l_\mu, d_i, m_\mu, \sigma_{i\mu} > 0 \text{ and } \gamma_{i\mu}, \alpha_{\mu i} \geq 0.} \quad (2.1)$$

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  $3N_R + 2N_S + 3N_R N_S$  parameters are constrained by the  $N_R + N_S$  equations (1.2). So if we choose  $2N_R + N_S + 3N_R N_S$  parameters, the remaining  $N_R + N_S$  are instantly

---

<sup>7</sup>It wouldn't make sense to say that species  $i$  eats resource  $\mu$  with efficiency 0, since this is equivalent to species  $i$  not eating resource  $\mu$ , 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  $m_\mu$ . 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:

$$\begin{cases} d_i = \sum_{\nu} (\sigma_{i\nu} \gamma_{i\nu} R_{\nu}^* - \alpha_{\nu i}) > 0 \quad \forall i = 1, \dots, N_S \\ m_\mu = \frac{l_\mu - \sum_j (\gamma_{j\mu} R_{\mu}^* - \alpha_{\mu j}) S_j^*}{R_{\mu}^*} > 0 \quad \forall \mu = 1, \dots, N_R \end{cases} \quad \begin{matrix} (2.2a) \\ (2.2b) \end{matrix}$$

In addition to these constraints, any feasible system should conserve biomass *at equilibrium*<sup>8</sup>; 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 :

$$\sum_{\nu} (1 - \sigma_{i\nu}) \gamma_{i\nu} R_{\nu}^* \geq \sum_{\nu} \alpha_{\nu i} \quad \forall i = 1, \dots, N_S. \quad (2.3)$$

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*  $\mathfrak{F} : \mathcal{P} \rightarrow \{0, 1\}$ , which takes a parameter set as an input and tells you whether this parameter set is feasible or not:

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

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 meta-parameter 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*  $\mathcal{F} : \mathcal{M} \rightarrow [0, 1] \times \mathcal{B}_{N_R \times N_S}$  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 :

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

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

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

<sup>8</sup>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 be made simpler. Feasibility conditions discussed above tell us that we may choose six metaparameters<sup>9</sup>:  $\gamma_0, \alpha_0, l_0, \sigma_0, S_0$  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<sup>10</sup>:

$$\langle R_\mu \rangle = R_0 = 1. \quad (2.7)$$

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

$$\langle l_\mu \rangle = l_0 = 1. \quad (2.8)$$

After this manipulation, the number of metaparameters is reduced from six to four : only  $\gamma_0, S_0, \alpha_0$  and  $\sigma_0$  remain.

For the sake of simplicity, we keep the same  $\sigma_0$  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:  $\alpha_0, \gamma_0$  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.* :

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

It is clear that  $\mathcal{F}_0^{G,A} = \mathcal{M} \forall G$  and  $\text{Vol}(\mathcal{F}_x^{G,A}) \leq \text{Vol}(\mathcal{F}_y^{G,A}) \forall x > y, G$ . We can similarly define for a set  $S = \{(G_1, A_1), (G_2, A_2), \dots, (G_N, A_N)\}$  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:

$$\mathcal{F}_x^S \equiv \bigcap_{(G,A) \in S} \mathcal{F}_x^{G,A}. \quad (2.10)$$

<sup>9</sup>Indeed, we saw that  $d_i$  and  $m_\mu$  are set by the other parameters, so we cannot freely choose  $d_0$  and  $m_0$ .

<sup>10</sup>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  $\gamma$  against  $\alpha$  instead of the more complicated  $\gamma R^*$  versus  $\alpha$ .

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 :

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

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 :

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

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:

$$\sum_{\nu} (1 - \sigma_{i\nu}) \gamma_{i\nu} R_{\nu}^* \geq \sum_{\nu} \alpha_{\nu i} \quad \forall i = 1, \dots, N_S. \quad (2.13)$$

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

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

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

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

Similarly,

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

Energy conservation Eq.(2.3) is then equivalent to

$$\deg(A, i) \alpha_0 \lesssim \deg(G, i) (1 - \sigma_0) R_0 \gamma_0 \quad \forall i = 1, \dots, N_S \quad (2.17)$$

Since  $\deg(G, i) > 0$ , we have<sup>11</sup>:

$$\frac{\deg(A, i)}{\deg(G, i)} \alpha_0 \lesssim (1 - \sigma_0) R_0 \gamma_0 \quad \forall i = 1, \dots, N_S \quad (2.18)$$

<sup>11</sup>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 :

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

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 :

$$\sum_{\mu} \sigma_{i\mu} \gamma_{i\mu} R_{\mu}^* > \sum_{\mu} \alpha_{\mu i} \quad (2.20)$$

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

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

Also, the resources diffusion rates  $m_{\nu}$  need to be positive:

$$l_{\nu} + \sum_j \alpha_{\nu j} S_j^* > \sum_j \gamma_{j\nu} R_{\nu}^* S_j^* \quad \forall \nu = 1, \dots, N_R \quad (2.22)$$

Which is equivalent to

$$l_0 + \deg(A, \nu) \alpha_0 S_0 \gtrsim \deg(G, \nu) \gamma_0 R_0 S_0 \quad \forall \nu \quad (2.23)$$

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

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

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  $\gamma$ '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  $\alpha$  (not sure about this for the last one).

### Combining both conditions

The two upper bounds Eqs. (2.19)-(2.21) on  $\alpha_0$  can be combined in a single inequality :

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

<sup>12</sup>Similarly to a previous footnote, we require that every resource  $\nu$  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  $\deg(A, i) \geq 1$ . Also, the largest value  $\deg(G, i)$  can get (for any  $i$ ) is  $N_R$ . Hence,

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

and we can find the largest allowed theoretical non-zero  $\alpha_0$  :

$$\alpha_0 \lesssim \min(1 - \sigma_0, \sigma_0) \gamma_0 R_0 N_R. \quad (2.27)$$

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

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



## 2.2 Dynamical stability

**TO DO:** change RC and write that it is expected to grow as the connectance of A grows, how does the proba(dyn | feas) make sense since I only consider points that are fully feasible (maybe put somewhere in appendix something like "in practice we only chose points such that F=1 for the study of D")? 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  $R_\mu^*$  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  $\{R_\mu^*, S_i^*\}$  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

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

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 :

$$J^* \equiv J(t_0), \quad (2.30)$$

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:

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

where  $\delta$  is the ubiquitously occurring Kronecker delta symbol defined as:

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

$J^*$  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:

$$\sum_\nu \sigma_{i\nu} \gamma_{i\nu} R_\nu^* - d_i - \sum_\nu \alpha_{\nu i} = 0, \quad (2.33)$$

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:

$$J^* = \begin{pmatrix} -\Delta & \Gamma \\ B & 0 \end{pmatrix}, \quad (2.34)$$

where

- $\Delta_{\mu\nu} = \text{diag}(m_\mu + \sum_j \gamma_{j\mu} S_j^*) = \text{diag}\left(\frac{l_\mu + \sum_j \alpha_{\mu j} S_j^*}{R_\mu^*}\right)$  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.
- $B_{i\nu} = \sigma_{i\nu} \gamma_{i\nu} S_i^*$  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 region $\mathcal{D}_{L,x}^{G,A}$

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

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

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

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

We also define the *metaparameters set local dynamical stability function*  $\mathcal{D}_L : \mathcal{M} \times \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S} \rightarrow [0, 1]$  which tells you, given a set of metaparameters  $m \in \mathcal{M}$  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:

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

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

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

Clearly,  $\mathcal{D}_{L,0}^{G,A} = \mathcal{M}$ ,  $\text{Vol}(\mathcal{D}_{L,x}^{G,A}) \leq \text{Vol}(\mathcal{D}_{L,y}^{G,A}) \forall x \geq y$ , 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  $S = \{(G_1, A_1) \dots, (G_N, A_N)\}$  their common  $x$  lds-region  $\mathcal{D}_{L,x}^S$ :

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

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:

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

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

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

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:

$$\begin{cases} R_\mu^* \rightarrow R_\mu(t_0) \equiv R_\mu^* (1 + \Delta_D \nu_\mu), \\ S_i^* \rightarrow S_i(t_0) \equiv S_i^* (1 + \Delta_D \nu_i), \end{cases} \quad (2.42)$$

$$\quad (2.43)$$

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  $\Delta_D$  to be non-infinitesimal. The question we will ask is precisely how big  $\Delta_D$  can get.

A certain number of quantities, that all depend on the perturbation  $\Delta_D$ , 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  $\alpha$  between two equilibria: this quantifies how close the old and new equilibria are.  $\alpha$  is defined through its standard scalar product formula:

$$\cos(\alpha) \equiv \frac{\sum_{\mu} R_{\mu}^* \tilde{R}_{\mu}^* + \sum_j S_j^* \tilde{S}_j^*}{\sqrt{\sum_{\mu} (R_{\mu}^*)^2 + \sum_i (S_i^*)^2} \sqrt{\sum_{\mu} (\tilde{R}_{\mu}^*)^2 + \sum_i (\tilde{S}_i^*)^2}}. \quad (2.44)$$

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  $\lambda_1$ , which is defined by the following property:

$$\boxed{\forall \lambda \in \sigma(J^*), \operatorname{Re}(\lambda) \leq \operatorname{Re}(\lambda_1)}, \quad (2.45)$$

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  $\lambda_1$  governs the local stability of the system at equilibrium **add source**. There are three cases:

- $\operatorname{Re}(\lambda_1) < 0$ : any perturbation on the abundances is exponentially suppressed. The system is stable.
- $\operatorname{Re}(\lambda_1) > 0$ : any perturbation on the abundances is exponentially amplified. The system is unstable.
- $\operatorname{Re}(\lambda_1) = 0$ : 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  $\text{Re}(\lambda_1)$ , 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:

$$\det(J^* - \lambda) = 0. \quad (2.46)$$

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

$$\det \begin{pmatrix} -\Delta - \lambda & \Gamma \\ \mathbf{B} & 0 - \lambda \end{pmatrix} = 0 \quad (2.47)$$

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

$$\lambda \neq 0. \quad (2.48)$$

Indeed<sup>13</sup> a non-zero  $\lambda$  implies

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

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]:

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

Hence Eq. (2.47) becomes:

$$\det(\lambda^2 \mathbb{1}_{N_R} + \Delta \lambda - \Gamma \mathbf{B}) = 0. \quad (2.51)$$

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  $\Gamma \mathbf{B}$   $N_R$ -dimensional square matrix, which is given by:

$$(\Gamma \mathbf{B})_{\mu\nu} = \sum_i \Gamma_{\mu i} \mathbf{B}_{i\nu} = \sum_i (\alpha_{\mu i} - \gamma_{i\mu} R_\mu^*) \sigma_{i\nu} \gamma_{i\nu} S_i^*. \quad (2.52)$$

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.

<sup>13</sup>Appendix 5.2.3 elaborates on when that condition is fulfilled.

**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:

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

In a more mathematical language:

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

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:*

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

*then  $\operatorname{Re}(\lambda) < 0 \forall \lambda \in \sigma(A)$ .*

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

$$|\lambda - A_{kk}| \leq \sum_{j \neq k} |A_{kj}|. \quad (2.56)$$

We now use the complex identity:

$$|\lambda - A_{kk}| \geq \operatorname{Re}(\lambda - A_{kk}) = \operatorname{Re}(\lambda) - \operatorname{Re}(A_{kk}). \quad (2.57)$$

Equation (2.55) implies:

$$\sum_{j \neq k} |A_{kj}| < -\operatorname{Re}(A_{kk}). \quad (2.58)$$

Combining the two previous inequalities yields:

$$\operatorname{Re}(\lambda) - \operatorname{Re}(A_{kk}) \leq |\lambda - A_{kk}| \leq \sum_{j \neq k} |A_{kj}| < -\operatorname{Re}(A_{kk}). \quad (2.59)$$

Comparing the RHS and LHS of this inequality yields:

$$\operatorname{Re}(\lambda) < 0. \quad (2.60)$$

□

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  $\lambda_1$ . 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:

$$D_\mu^R \equiv \left\{ z \in \mathbb{C} : |z + \Delta_\mu| \leq \sum_j |\Gamma_{\mu j}| \right\} \quad \forall \mu = 1, \dots, N_R, \quad (2.61)$$

and the “consumers” discs:

$$D_i^C \equiv \left\{ z \in \mathbb{C} : |z| \leq \sum_\nu |B_{i\nu}| \right\} \quad \forall i = 1, \dots, N_S. \quad (2.62)$$

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

$$|\lambda| \leq \sum_\nu |B_{i^* \nu}| \quad (2.63)$$

or

$$|\lambda + \Delta_{\mu^*}| \leq \sum_j |\Gamma_{\mu^* j}| \quad (2.64)$$

The triangle inequality implies:

$$|\lambda| \leq |\lambda + \Delta_{\mu^*}| + |-\Delta_{\mu^*}| \leq \sum_j |\Gamma_{\mu^* j}| + |-\Delta_{\mu^*}| = \sum_j |\Gamma_{\mu^* j}| + \Delta_{\mu^*}. \quad (2.65)$$

The only way both Eq. (2.63) and (2.65) are satisfied for all eigenvalues, and especially the one with the highest real part  $\lambda_1$  is if they are bound by the maximum of both RHS of these equations. More precisely:

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

where we defined the critical radius  $R_C$  as:

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

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:

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

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

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

In that regime we then have:

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

Similarly we find

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

such that  $R_C$  can be estimated roughly as:

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

### 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*  $\text{Re}(\lambda_1) < 0$ , 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<sup>14</sup>

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

with

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

or, component-wise:

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

The strategy is to assume we are in an unstable regime, i.e. there exists at least one  $\lambda \in \sigma(J^*)$  with  $\text{Re}(\lambda) \geq 0$  that satisfies Eq. (2.51) and such that  $\text{Re}(\lambda) > 0$ . By Eq. (2.73),  $\lambda$  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  $\text{Re}(\lambda) \leq 0$ . As this is a contradiction to the hypothesis that the regime is unstable, we must conclude that the regime is stable<sup>15</sup>

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  $\lambda$ . Thanks to the help of the two following theorems, we found such a regime, called *low intra-resource interaction* or LRI.

<sup>14</sup>We can do this because since  $m_{\mu} > 0$ , we know  $\Delta$  will always be invertible.

<sup>15</sup>Indeed, Eq. (2.51) assumes already that either  $\text{Re}(\lambda_1) > 0$  or  $\text{Re}(\lambda_1) < 0$ .



**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

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

are verified, then  $p$  is dynamically stable.

*Proof.* We assume

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

This implies:

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

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

$$(\Gamma B)_{\mu\mu} + \text{Im}(\lambda)^2 < - \sum_{\nu \neq \mu} |(\Gamma B)_{\mu\nu}| \quad \forall \mu. \quad (2.79)$$

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

$$\text{Im}(c)^2 \geq -\text{Re}(c^2) \quad \forall c \in \mathbb{C}. \quad (2.80)$$

Using this result and dividing Eq.(2.79) by  $\Delta_\mu$ , we get:

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

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

$$\text{Re}(S_{\mu\mu}) + \sum_{\nu \neq \mu} |S_{\mu\nu}| < 0 \quad \forall \mu. \quad (2.82)$$

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

Theorem 1 is a strong statement but it asks a lot on the parameters set, namely  $\Gamma B$  must have diagonal elements that are “very negative”, which imposes severe conditions especially on the  $\alpha$  and  $\gamma$  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.

---

<sup>16</sup>That step is valid because  $\Delta_\mu > 0$ ,  $\forall \mu$ .

### 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*

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

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

### 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  $(\Gamma B)_{\mu\nu}$ , 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:

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

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

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

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  $\mu$  and consumes resource  $\nu$ . Hence  $O_{\mu\nu}$  essentially tells how many species effectively “link” resource  $\mu$  to resource  $\nu$  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  $\mu$  and  $\nu$ .

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

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

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

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

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  $\alpha_0$ . Indeed, since the RHS of the previous equation is negative, we need:

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

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  $\alpha_0$  for feasible systems:

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

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  $\mu$ , 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  $\Delta E$  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  $\Delta E$  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  $\Delta E$  between the new proposed  $A'$  and the old  $A$  is computed by assigning an energy  $E$  to both  $A'$  and  $A$  and subtracting them:

$$\Delta E \equiv E(A', G) - E(A, G). \quad (2.92)$$

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):

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

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  $\alpha_0$  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  $\alpha_0$  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.*

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

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  $\tilde{q}_{i\mu}$ :

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

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

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

and  $\sigma_q$  is the standard deviation of  $q$ , again element-wise:

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

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  $\sigma_q$  multiplied by a factor of magnitude  $\sim 1$ . Indeed the  $\tilde{q}_{i\mu}$  are not large since they follow the two equalities [27]:

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

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  $\Omega$ , we can write the following approximation in the case  $N_R, N_S \gg 1$ :

$$\langle q \rangle \approx \langle \Omega \rangle \equiv q_0. \quad (2.99)$$

We can then rewrite the free parameters of our model<sup>17</sup>:

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

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 :

$$\begin{cases} \frac{l_\mu + \sum_j \alpha_{\mu j} S_j^*}{R_\mu^*} = \frac{l_0 + N_S \alpha_0 S_0 + \sigma_l \tilde{l}_\mu + \sigma_\alpha S_0 \sum_j \tilde{a}_{\mu j} + \sigma_\alpha \sigma_S \sum_j \tilde{a}_{\mu j} \tilde{s}_j}{R_0 + \sigma_R \tilde{r}_\mu} & (2.101) \end{cases}$$

$$\begin{cases} -\gamma_{j\mu} R_\mu^* + \alpha_{\mu j} = -\gamma_0 R_0 + \alpha_0 + \sigma_\gamma R_0 \tilde{g}_{j\mu} + \sigma_R \gamma_0 \tilde{r}_\mu + \sigma_\alpha \tilde{a}_{\mu j} + \sigma_\gamma \sigma_R \tilde{g}_{j\mu} \tilde{r}_\mu & (2.102) \end{cases}$$

$$\begin{cases} \sigma_{i\nu} \gamma_{i\nu} S_i^* = \sigma_0 \gamma_0 S_0 + \sigma_\sigma \gamma_0 S_0 \tilde{\sigma}_{i\nu} + \sigma_\gamma \sigma_0 S_0 \tilde{g}_{i\nu} + \sigma_S \sigma_0 \gamma_0 \tilde{s}_i + \sigma_\sigma \sigma_\gamma S_0 \tilde{\sigma}_{i\nu} \tilde{g}_{i\nu} \\ \quad + \sigma_\sigma \sigma_S \gamma_0 \tilde{\sigma}_{i\nu} \tilde{s}_i + \sigma_\gamma \sigma_S \sigma_0 \tilde{g}_{i\nu} \tilde{s}_i + \sigma_\sigma \sigma_\gamma \sigma_S \tilde{\sigma}_{i\nu} \tilde{g}_{i\nu} \tilde{s}_i & (2.103) \end{cases}$$

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

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

The previous relations then become :

$$\begin{cases} \frac{l_\mu + \sum_j \alpha_{\mu j} S_j^*}{R_\mu^*} = \frac{l_0 (1 + \epsilon_l \tilde{l}_\mu) + \alpha_0 S_0 (N_S + \epsilon_\alpha \sum_j \tilde{a}_{\mu j} + \epsilon_\alpha \epsilon_S \sum_j \tilde{a}_{\mu j} \tilde{s}_j)}{R_0 (1 + \epsilon_R \tilde{r}_\mu)} & (2.105) \end{cases}$$

$$\begin{cases} -\gamma_{j\mu} R_\mu^* + \alpha_{\mu j} = -\gamma_0 R_0 (1 + \epsilon_\gamma \tilde{g}_{j\mu} + \epsilon_R \tilde{r}_\mu + \epsilon_\gamma \epsilon_R \tilde{g}_{j\mu} \tilde{r}_\mu) + \alpha_0 (1 + \epsilon_\alpha \tilde{a}_{\mu j}) & (2.106) \end{cases}$$

$$\begin{cases} \sigma_{i\nu} \gamma_{i\nu} S_i^* = \sigma_0 \gamma_0 S_0 (1 + \epsilon_\sigma \tilde{\sigma}_{i\nu} + \epsilon_\gamma \tilde{g}_{i\nu} + \epsilon_S \tilde{s}_i + \epsilon_\sigma \epsilon_\gamma \tilde{\sigma}_{i\nu} \tilde{g}_{i\nu} + \epsilon_\sigma \epsilon_S \tilde{\sigma}_{i\nu} \tilde{s}_i \\ \quad + \epsilon_\gamma \epsilon_S \tilde{g}_{i\nu} \tilde{s}_i + \epsilon_\sigma \epsilon_\gamma \epsilon_S \tilde{\sigma}_{i\nu} \tilde{g}_{i\nu} \tilde{s}_i) & (2.107) \end{cases}$$

### Standard deviation expansion at first order

We assume the relative standard deviations are small:

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

<sup>17</sup>This works with  $\gamma$  and  $\alpha$  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 previous equations can be rewritten as:

$$\left\{ \begin{array}{l} \frac{l_\mu + \sum_j \alpha_{\mu j} S_j^*}{R_\mu^*} = \frac{l_0 + N_S \alpha_0 S_0}{R_0} - \epsilon_R \frac{l_0 + N_S \alpha_0 S_0}{R_0} \tilde{r}_\mu + \epsilon_l \frac{l_0}{R_0} \tilde{l}_\mu + \epsilon_\alpha \frac{\alpha_0 S_0}{R_0} \sum_j \tilde{\alpha}_{\mu j} \end{array} \right. \quad (2.109)$$

$$\left\{ \begin{array}{l} -\gamma_{j\mu} R_\mu^* + \alpha_{\mu j} = -\gamma_0 R_0 + \alpha_0 - \epsilon_\gamma \gamma_0 R_0 \tilde{\gamma}_{j\mu} - \epsilon_R \gamma_0 R_0 \tilde{r}_\mu + \epsilon_\alpha \alpha_0 \tilde{\alpha}_{\mu j} \\ \sigma_{i\nu} \gamma_{i\nu} S_i^* = \sigma_0 \gamma_0 S_0 (1 + \epsilon_\sigma \tilde{\sigma}_{i\nu} + \epsilon_\gamma \tilde{\gamma}_{i\nu} + \epsilon_S \tilde{s}_i) \end{array} \right. \quad (2.110)$$

$$\sigma_{i\nu} \gamma_{i\nu} S_i^* = \sigma_0 \gamma_0 S_0 (1 + \epsilon_\sigma \tilde{\sigma}_{i\nu} + \epsilon_\gamma \tilde{\gamma}_{i\nu} + \epsilon_S \tilde{s}_i) \quad (2.111)$$

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 :

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

This allows us to rewrite the jacobian at equilibrium as:

$$J^* = J_0 + \epsilon \tilde{J}, \quad (2.113)$$

with

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

where

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

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

$$(B_0)_{i\nu} \equiv B_0 = \sigma_0 \gamma_0 S_0 \quad (2.117)$$

and

$$\tilde{J} = \begin{pmatrix} \tilde{\Delta} & \tilde{\Gamma} \\ \tilde{B} & 0 \end{pmatrix} \quad (2.118)$$

with

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

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

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

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

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

where  $\operatorname{adj}(\dots)$  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:

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

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

$$\det(\lambda^2 \mathbb{1}_{N_R} + \Delta_0 \lambda - \Gamma_0 B_0) = 0 \quad (2.124)$$

Component-wise, we have:

$$(\lambda^2 \mathbb{1}_{N_R} + \Delta_0 \lambda - \Gamma_0 B_0)_{\mu\nu} = (\lambda^2 + \Delta_0 \lambda) \delta_{\mu\nu} - \Gamma_0 B_0. \quad (2.125)$$

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:

$$(\lambda + \Delta_0)^{N_R-1} (\lambda^2 + \Delta_0 \lambda - N_R \Gamma_0 B_0) = 0. \quad (2.126)$$

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:

$$\lambda^2 + \Delta_0 \lambda - N_R \Gamma_0 B_0 = 0. \quad (2.127)$$

In the end, the spectrum is given by:

- if  $\Gamma_0 < -\frac{\Delta_0^2}{4N_R B_0}$  :

$$\sigma(J_0) = \left\{ 0, \dots, 0, -\Delta_0, \dots, -\Delta_0, -\frac{\Delta_0}{2} \left( 1 \pm i \sqrt{1 + \frac{4N_R \Gamma_0 B_0}{\Delta_0^2}} \right) \right\} \quad (2.128)$$

- if  $\Gamma_0 = -\frac{\Delta_0^2}{4N_R B_0}$  :

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

- if  $\Gamma_0 > -\frac{\Delta_0^2}{4N_R B_0}$  :

$$\sigma(J_0) = \left\{ 0, \dots, 0, -\Delta_0, \dots, -\Delta_0, -\frac{\Delta_0}{2} \left( 1 \pm \sqrt{1 + \frac{4N_R \Gamma_0 B_0}{\Delta_0^2}} \right) \right\} \quad (2.130)$$

It then becomes clear that the system is dynamically unstable if and only if  $\frac{4N_R \Gamma_0 B_0}{\Delta_0^2} > 0$ . Because  $N_R, B_0 > 0$ , we get the condition:

The non-variance system is dynamically unstable  $\iff \Gamma_0 > 0$ .

(2.131)



If  $\Gamma_0 \leq 0$ , 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 :

$$\alpha_0 \lesssim \min(1 - \sigma_0, \sigma_0) \gamma_0 R_0 \iff \Gamma_0 \lesssim [\min(1 - \sigma_0, \sigma_0) - 1] \gamma_0 R_0 < 0, \quad (2.132)$$

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  $4N_R\Gamma_0\mathbf{B}_0 + \Delta_0^2$ . 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  $4N_R\Gamma_0\mathbf{B}_0 + \Delta_0^2$  is small, *i.e.*

$$4N_R\Gamma_0\mathbf{B}_0 + \Delta_0^2 \ll 1 \iff 4N_R\sigma_0\gamma_0S_0(\alpha_0 - \gamma_0R_0) + \frac{l_0^2}{R_0^2} + \frac{2N_S\alpha_0S_0l_0}{R_0^2} + \frac{N_S^2\alpha_0^2S_0^2}{R_0^2} \ll 1. \quad (2.133)$$

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  $l_\mu$ , 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  $\Delta_S$ , if under the transformation

$$l_\mu \rightarrow \hat{l}_\mu \equiv l_\mu (1 + \Delta_S \nu_\mu) \quad (2.134)$$

the transformed set of parameters  $\{\hat{l}_\mu, m_\mu, \gamma_{i\mu}, \alpha_{\mu i}, \sigma_{i\mu}, d_i\}$  gives rise under time evolution to a positive valued-equilibrium  $\{\hat{R}_\mu^*, \hat{S}_i^*\}$ . In the equation above,  $\nu_\mu$  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*  $\mathfrak{S} : [0, 1] \times \mathcal{P} \rightarrow \{0, 1\}$  in the following way  $\forall \Delta_S \in [0, 1], p \in \mathcal{P}$ :

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

For a fixed  $p$ , we expect  $\mathfrak{S}(\Delta_S, p)$  to behave as a step function of  $\Delta_S$  : 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  $\Delta_S$ . Mathematically,  $\mathcal{S} : [0, 1] \times \mathcal{M} \times \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S} \rightarrow [0, 1]$  is defined  $\forall \Delta_S \in [0, 1], m \in \mathcal{M}, B = (G, A) \in \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S}$  :

$$\mathcal{S}(\Delta_S, m, B) = \text{Probability} \{ \mathfrak{S}(\Delta_S, \mathcal{A}(m, B)) = 1 \} \quad (2.136)$$

Because we expect a step-like drop of  $\mathfrak{S}$  as  $\Delta_S$  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 :

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

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  $\Delta_S^*$  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  $\Delta_S$  and it is clear that

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

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 [5.2.9](#)).



### 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<sup>18</sup> 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  $\gamma$ , but also to the syntrophy matrix  $\alpha$ . 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  $\kappa_G$  and nestedness (which, for the case of  $G$ , we call ecological overlap)  $\eta_G$ . 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"<sup>19</sup>:

You should talk at some point about the number of resources, G25 and G50 I think you call it later

- "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.

<sup>18</sup>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<sup>Is this true?</sup> 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].

<sup>19</sup>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<sup>[20]</sup> 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

**TO DO : explain why high  $\gamma_0$  and low  $S_0$  make sense biologically, say that for high  $\kappa$  and small  $\epsilon$ , "syntrophy impacts the existence least"** 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  $\alpha_0$ , what values  $(\gamma_0, S_0)$  can take such that  $m = (\gamma_0, S_0, \alpha_0)$  is in  $\mathcal{F}_1^{G,A}$ . for different values of alpha are compatible with full feasibility, i.e.  $m = \dots$

#### 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:

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

<sup>20</sup>Note that we will take a constant value  $\alpha_0$  (given in Methods [2.2.4](#)) and  $\gamma_0 = 0.2$ . A more thorough analysis should build the optimal LRI matrix *corresponding to each*  $(\gamma_0, \alpha_0)$ . That would take too much time which is why we decided to keep  $\gamma_0$  and  $\alpha_0$  constant.

Wouldn't be easier to say, everything else being fixed, increasing either the consumption rate, the resources at equilibrium, the species abundances at equilibrium and the number of resources being consumed will decrease feasibility  
input rate of resources also decreases feasibility. You could actually write the eq as  $\gamma_0 R_0 S_0 \max(\deg) \leq I_0$

### 3.1. FEASIBILITY

### CHAPTER 3. RESULTS

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

- Given a feasible system ...
  - For a fixed consumption interaction (i.e. constant  $\gamma_0$  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. that, in order to maintain feasibility, the external resources also increase
  - For a fixed consumption network  $G$ , resources abundance  $R_0$  and resource input rate  $I_0$ , a larger range of possible consumption rate  $\gamma_0$  is possible if the consumers abundance is decreased. Since there are fewer the consumers, they can individually eat more from the available resources. I like more the (also obvious) observation in which for fixed  $R_0$ , increasing  $\gamma_0$  requires increasing the input of external resources
  - For every metaparameter except  $\gamma_0$  fixed, arranging the shape of  $G$  such that few consumers eat from the same resources<sup>21</sup> 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  $(\gamma_0, S_0)$  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<sup>22</sup>  $\mathcal{F}((\gamma_0, S_0, 0), G, 0) = 0.5$ . Theoretically, that sharp transition between feasibility and infeasibility happens when both sides of the inequality (3.1) are equal, i.e. at  $\gamma_0 R_0 = I_0 / \max_{\nu} \{\deg(G, \nu)\} S_0$ . Hence we expect the boundary measured above to be well characterized by the curve  $S_0 = K \gamma_0^{-1}$  with  $K = I_0 / (R_0 \max_{\nu} \{\deg(G, \nu)\})$ . cool

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  $\Delta_K \equiv 1 - K_{\text{theoretical}} / K_{\text{fit}}$ . 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

<sup>21</sup>Indeed,  $\deg(G, \nu)$  is the number of species that consume resource  $\nu$ . Reducing the number of consumers that eat from the same species reduces  $\max \deg(G, \nu)$  in Equation (3.1). resources

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

I think that perhaps you were a little bit too strict with your comparison because the difference in K may be up to a 20% but here we are focusing on the "feasibility region" which is the whole area in the gamma0, S0 plot. If you compute the area enclosed by the theoretical and fitted curves (and the gamma0, S0 axes) and you compute the relative difference, it will be much smaller

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 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  $\alpha_0$ ? More precisely, how does the shape of  $\mathcal{F}_1^{G,A}$  change as a function of  $A$  and  $\alpha_0$ ?

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  $\alpha_0$  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<sub>1</sub>, LRI<sub>1</sub>, RS<sub>1</sub>). Concerning the question of  $\alpha_0$ , 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  $\alpha_0$  such that  $\mathcal{F}_1^{G,A} \neq \emptyset \forall (G, A) \in S_{25}$ , can be estimated with the help of Eq. (2.28):

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

I think it is important to remark here why L0 does not appear in the Eq as in the a=0 case, it is an upper bound fixed by choice of units, fair comparison...

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

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

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  $\alpha_0$  but also with the network structure of the consumption matrix  $G$ . How  $\mathcal{F}_1^{G,A}(\alpha_0)$  changes

The way in which F1



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  $d_F$  and  $\alpha_0$  change from the null case)?**

the fully feasible region shrinks towards high  $\gamma_0$  don't follow you here

We observe a general trend among all matrices: as syntrophy increases, the fully feasible region moves horizontally to the right towards a higher  $\gamma_0$ . This can be explained with LHS Eq. (2.28) which provides a lower bound to  $\gamma_0$  when  $\alpha_0 > 0$ . Note that taking into account  $\alpha_0 > 0$  does not bound  $S_0$ , such that at a fixed  $\gamma_0$ , 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**

Perhaps you could simply write LHS of eq 2.38 writing alpha/gamma explicitly?

conservation of biomass, make it explicit,

Because of that shift to the right,  $\mathcal{F}_1^{G,A}(\alpha_0)$  shrinks in size<sup>23</sup> 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  $\text{Vol}(\mathcal{F}_1^{G,A}(\alpha_0))$ , formally defined in Appendix 5.2.6, decays in an exponential-like fashion as  $\alpha_0$  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<sup>24</sup>:

the best coefficients satisfying

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

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<sup>25</sup>. 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. it will become unfeasible, i.e. at least one species will get extinct.

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  $\kappa_G$ ,  $d_F$  is increased

I do not agree in that it does not make a difference, a 10-20% is quite notable, and actually if you remove the outlier (which is how I would present it, a clear trend could be observed)

NIS?

<sup>23</sup>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.

<sup>24</sup>In practice, standard functions of the numpy and scipy Python libraries are used to perform that fitting procedure. this is a Method and actually technical, perhaps you should add a section (maybe suppl.) with these things

<sup>25</sup>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.

Like, all code was written in Python and C, libraries blabla for the plots blabla for fits

if the ecological overlap  $\eta_G$  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 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:

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

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  $\alpha_0$ , there are more feasible  $(\gamma_0, S_0)$  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  $(\gamma_0, S_0)$  that are fully feasible for all matrices. This means that the largest common fully feasible syntrophy respects  $7.8 \times 10^{-3} < \alpha_C^F(S_{25}) \leq 9.1 \times 10^{-3}$  for those three scenarios, which puts the estimate  $\alpha_C^F(S_{25}) \approx 0.01$  of Eq. (3.2) not far from reality. **reality?**

Without surprise,  $\text{Vol}(\mathcal{F}_1^{S_{25}}(\alpha_0))$  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  $\alpha_0$ ); 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  $\alpha_0$ , for **much** more pairs of common feasible  $(\gamma_0, S_0)$ . **Without any surprise as well** we observe the same shift of  $\mathcal{F}_1^{S_{25}}$  towards points with a high  $\gamma_0$  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. **To keep in mind for the discussion this message. In the discussion one of the things that are expected is your interpretation of the evolutionary processes that may lead to syntrophy, the fact that random structures are more feasible are actually good news, because it does not require any fine-tuning to improve feasibility**

with high uptake rates (but not arbitrarily large because are upper bounded by  $L_0$ )

many

low abundance

### 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

limiting case

<sup>26</sup>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].

It is controversial and unfortunately we do not say much on that. It is known that non-linear trends could bring regions of stability even with less resources

In the previous section I think you could have remarked that the feasibility with syntrophy is always smaller than without syntrophy, is it also the case here? An interesting plot for the future could be, for a fixed topology of  $G$ , changing NR and quantifying the volume variation, figure 3.12 is quite noisy

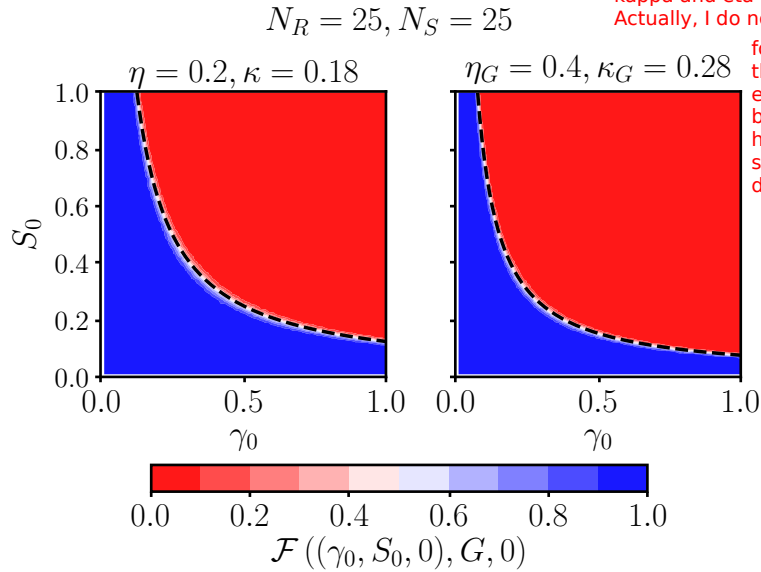
fruitful to investigate the behaviour of systems where the number of resources has been increased to  $N_R = 50$ .

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  $1.4 \times 10^{-2}$  to  $7.8 \times 10^{-3}$ .

Difficult to say by eye, I would refer to 3.12 in general, this actually may be suppl. figs. I actually do not know if these figures are the last ones

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 3.2.17 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**

I think it is important to note that changing NR makes two networks, even if they have the same kappa and eta values might not be comparable. Actually, I do not remember how I computed eta



for the NR=50 but if I did it considering the mean of eta(consumers) and eta(resources) the two networks could be quite different. Perhaps I should have created networks with the same eta(consumers) which is the dimension that do not change

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.

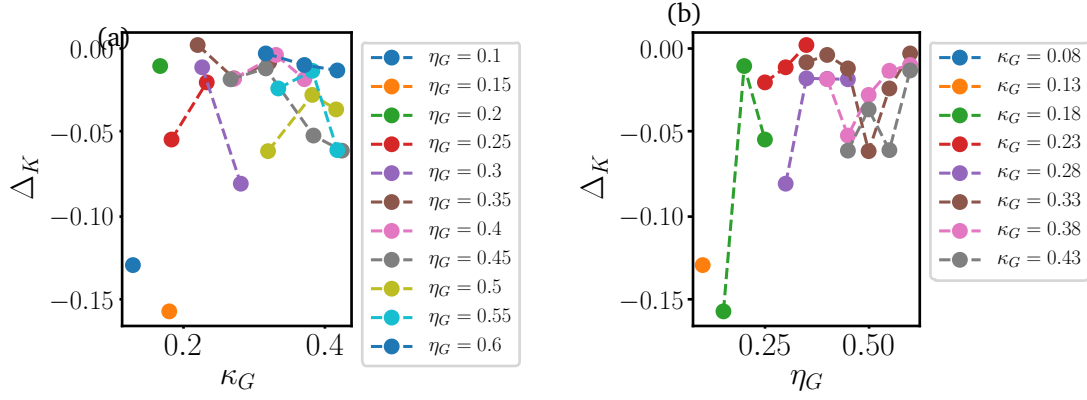


Figure 3.1.2: Relative error in the determination of the boundary of  $\mathcal{F}_1^{G,0}$  (a) varying connectance at fixed ecological overlap and (b) varying ecological overlap at fixed connectance. The theoretical prediction tends to overestimate the measured value. The larger the ecological overlap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

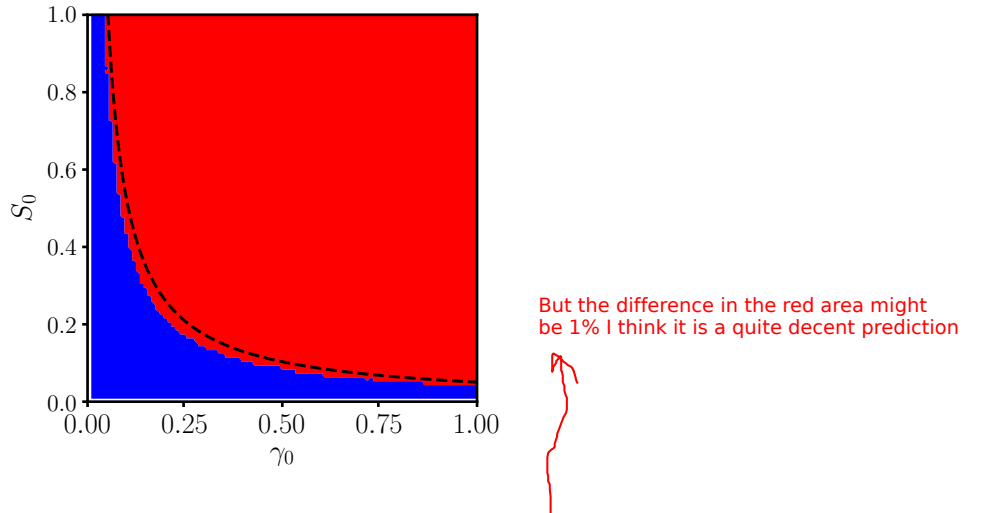


Figure 3.1.3: Plot of the common feasibility region. The blue area indicates the common feasibility volume, computed numerically, while the dashed line shows the analytical prediction. Although the match is not as good as before, the relative error is ~~only~~ of the order of 20%. The red part is the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

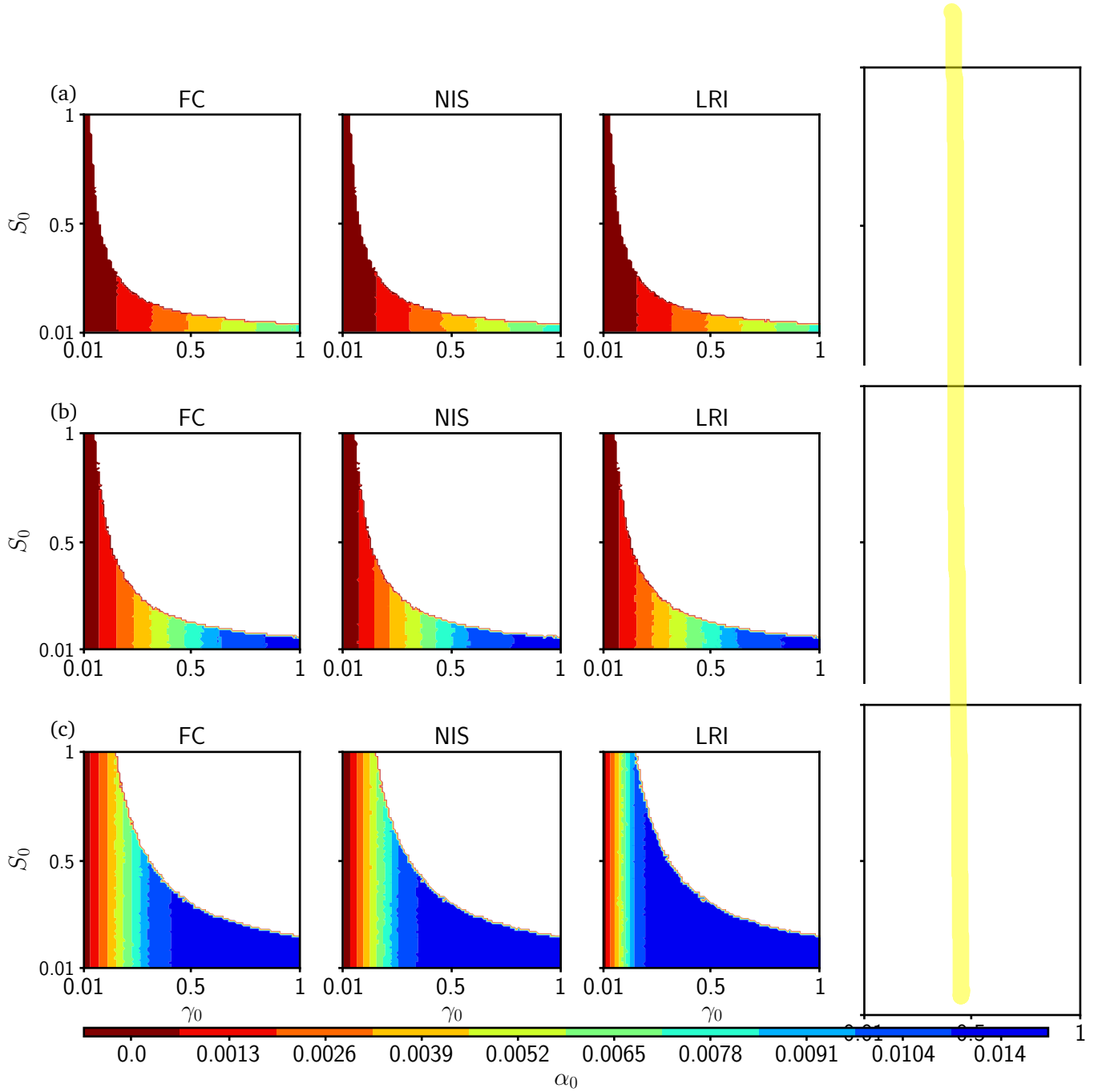


Figure 3.1.4: Fully feasible region in the  $(\gamma_0, S_0) \in [0, 1] \times [0, 1]$  unit square as a function of syntrophy for different consumption matrices  $G$ : (a)  $\eta_G = 0.6$ ,  $\kappa_G = 0.32$ , (b)  $\eta_G = 0.35$ ,  $\kappa_G = 0.22$  and (c)  $\eta_G = 0.15$ ,  $\kappa_G = 0.18$ . 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  $0 \leq \alpha_0 \leq 9.1 \times 10^{-3}$ . The size of the feasibility regions depend heavily on the topology of the matrix, which makes the problem far from trivial.

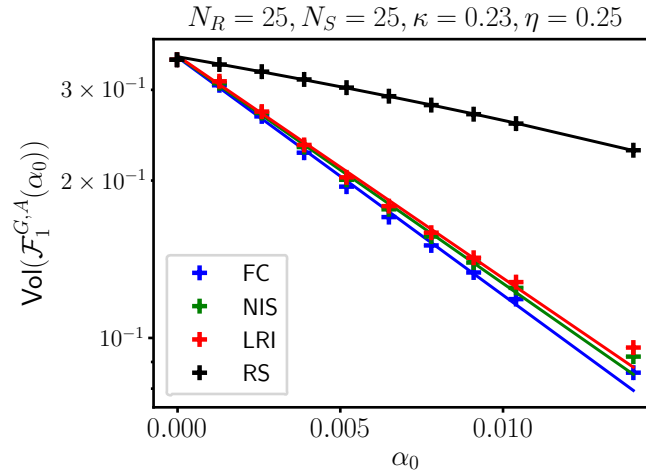


Figure 3.1.5: Decay of the volume of the fully feasible region  $\mathcal{F}_1^{G,A}(\alpha_0)$  for a matrix consumption  $G$  with ecological overlap  $\eta_G = 0.25$  and connectance  $\kappa_G = 0.23$  on a logarithmic scale. The solid lines represent the exponential fit explained in the main text.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

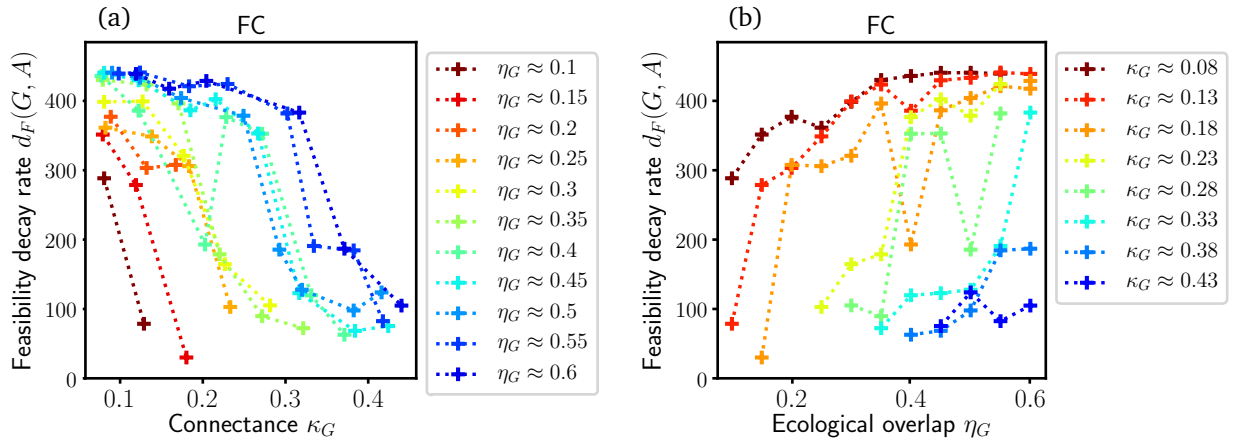


Figure 3.1.6: Feasibility decay rate  $d_F(G, A)$  for  $A$  fully connected and  $(G, A) \in S_{25}$ . (a)  $d_F$  as a function of the connectance of  $G$  for different fixed ecological overlaps and (b)  $d_F$  as a function of the ecological overlap  $\eta_G$  for fixed different connectances. A 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.

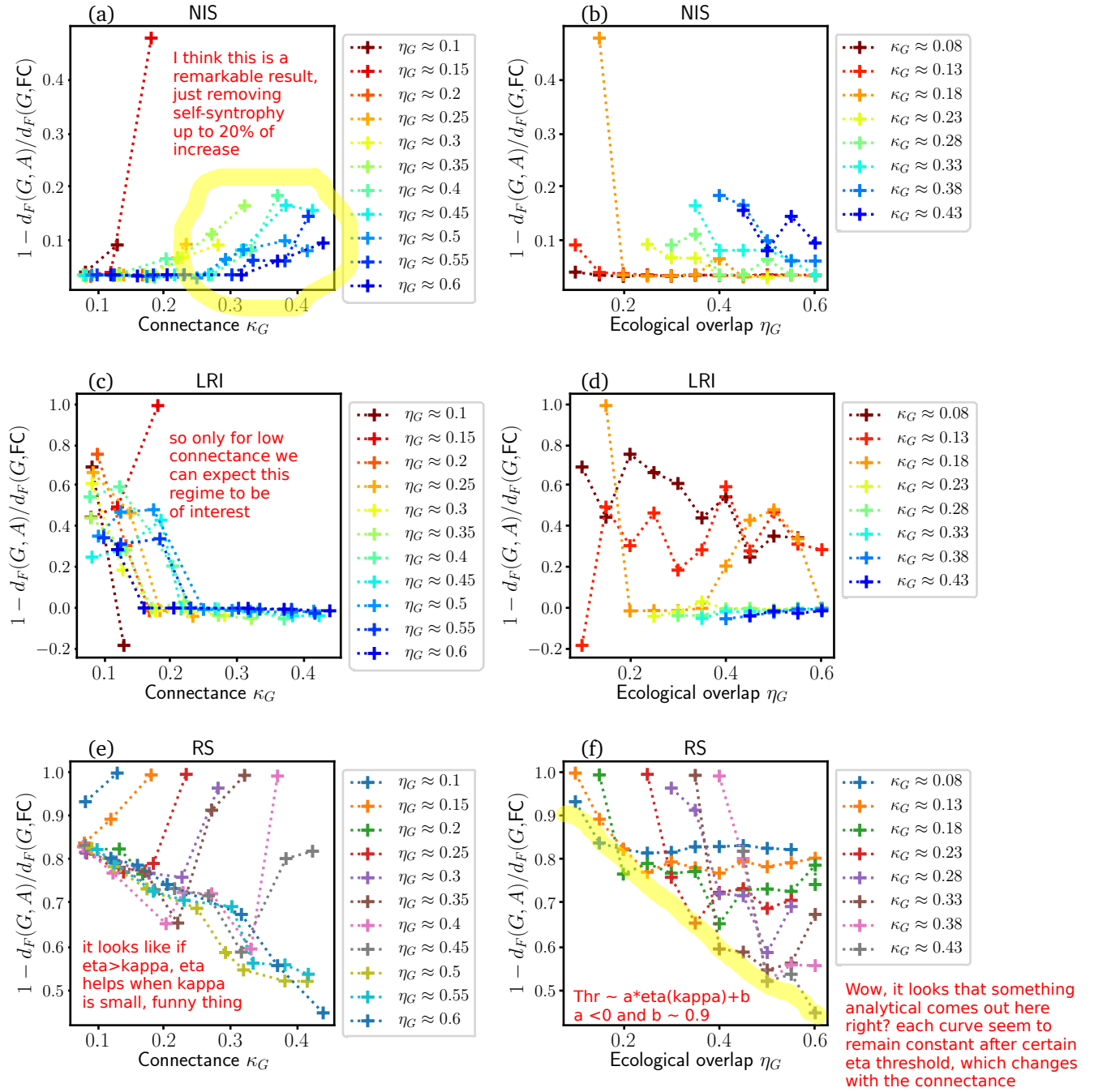


Figure 3.1.7: Relative difference of the feasibility decay rate for the considered  $A$  scenario compared to the FC null case (Fig. 3.1.6) for  $N_R = 25$  and  $N_S = 25$ . Plots on the first column (a)-(c)-(e) show how that quantity changes with connectance for a given ecological overlap, while plot 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.

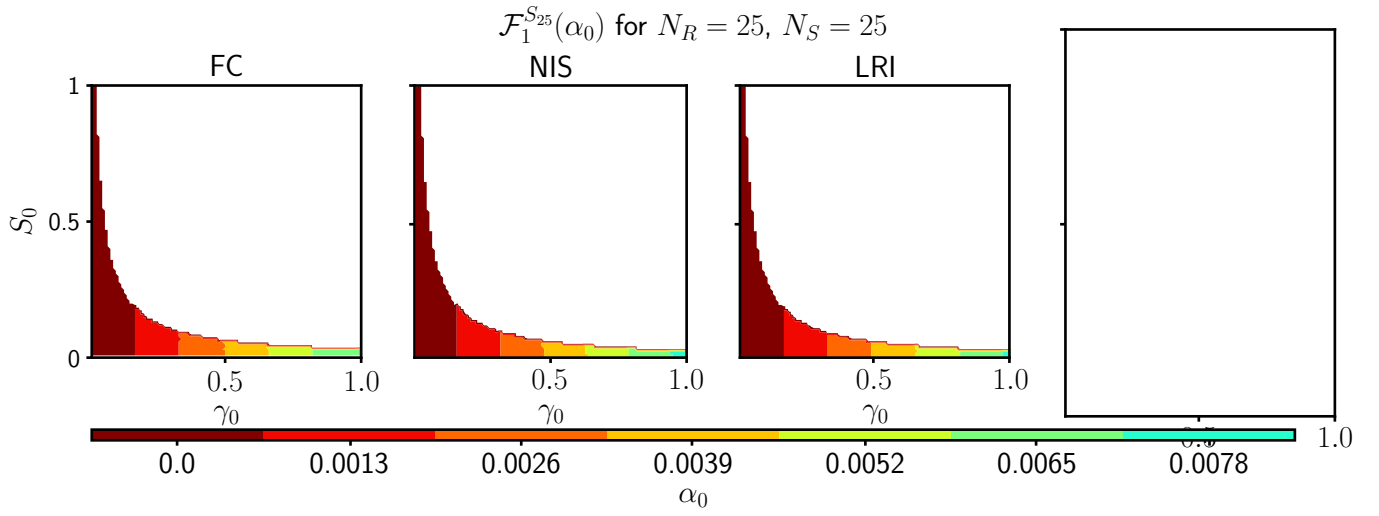


Figure 3.1.8: Surface plot of the fully feasible volume  $\mathcal{F}_1^{S_{25}}(\alpha_0)$ . The color bar on the side indicates the value of  $\alpha_0$  to which the surface corresponds. The white part of the plot corresponds to points that never are fully feasible. Note that even though it is not 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.

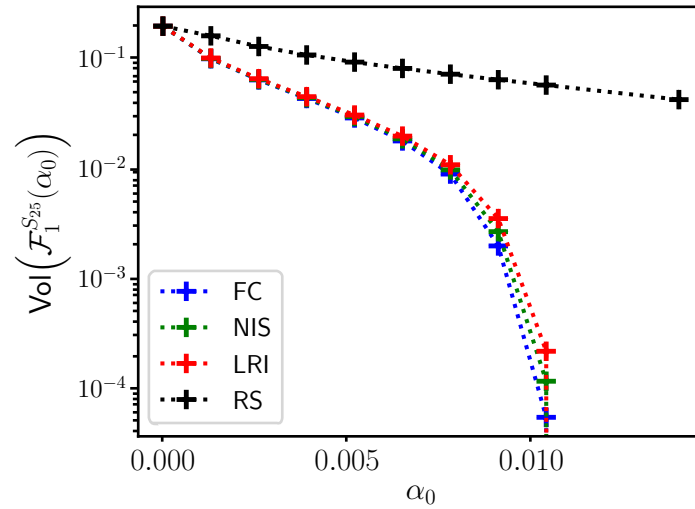


Figure 3.1.9: Volume of the common feasibility region  $\mathcal{F}_1^{S_{25}}(\alpha_0)$  as a function of syntrophic interaction strength  $\alpha_0$  (plotted on logarithmic scale). While the FC, NIS and LRI cases offer similar results, the RS scenario outperforms all of them. An exponential fit (Eq. 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.



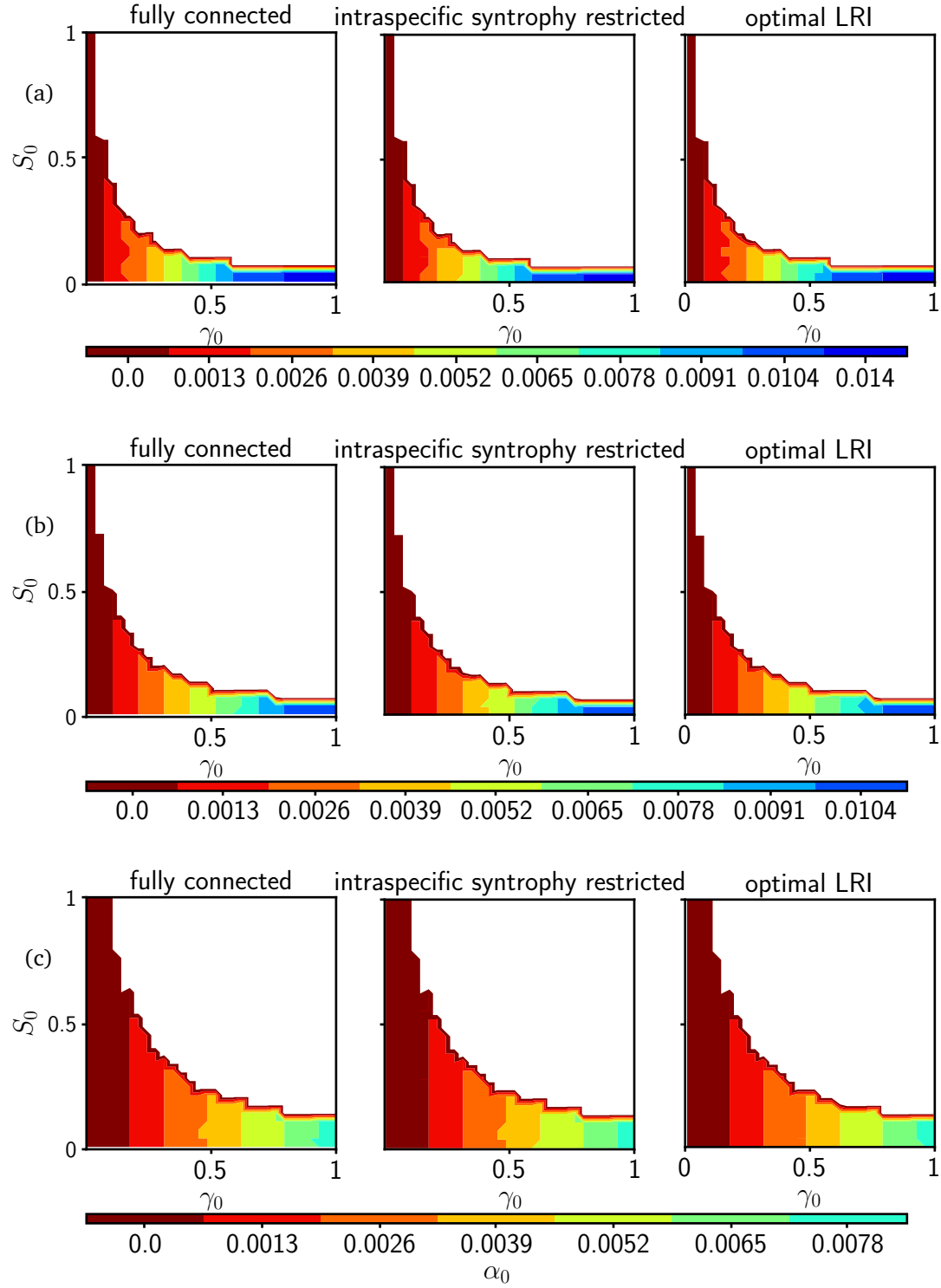


Figure 3.1.10: Surface colour plot of the fully feasible region  $\mathcal{F}_1^{G,A}(\alpha_0)$  as a function of the syntrophy  $\alpha_0$  for the case  $N_R = 50$ , with different structures of  $A$ : fully connected (left column), no intraspecific syntrophy (middle) and LRI matrix (right). The rows 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.

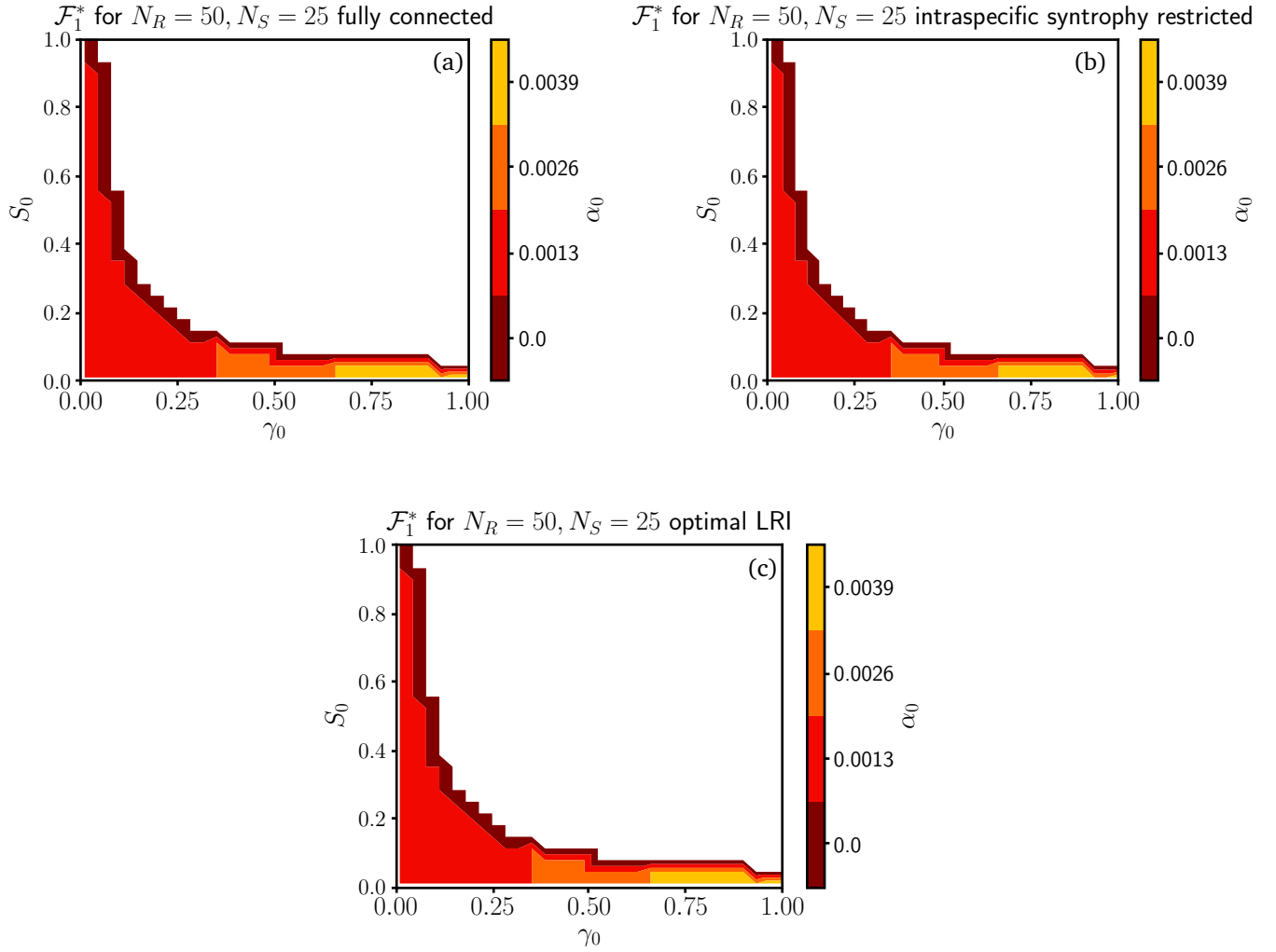
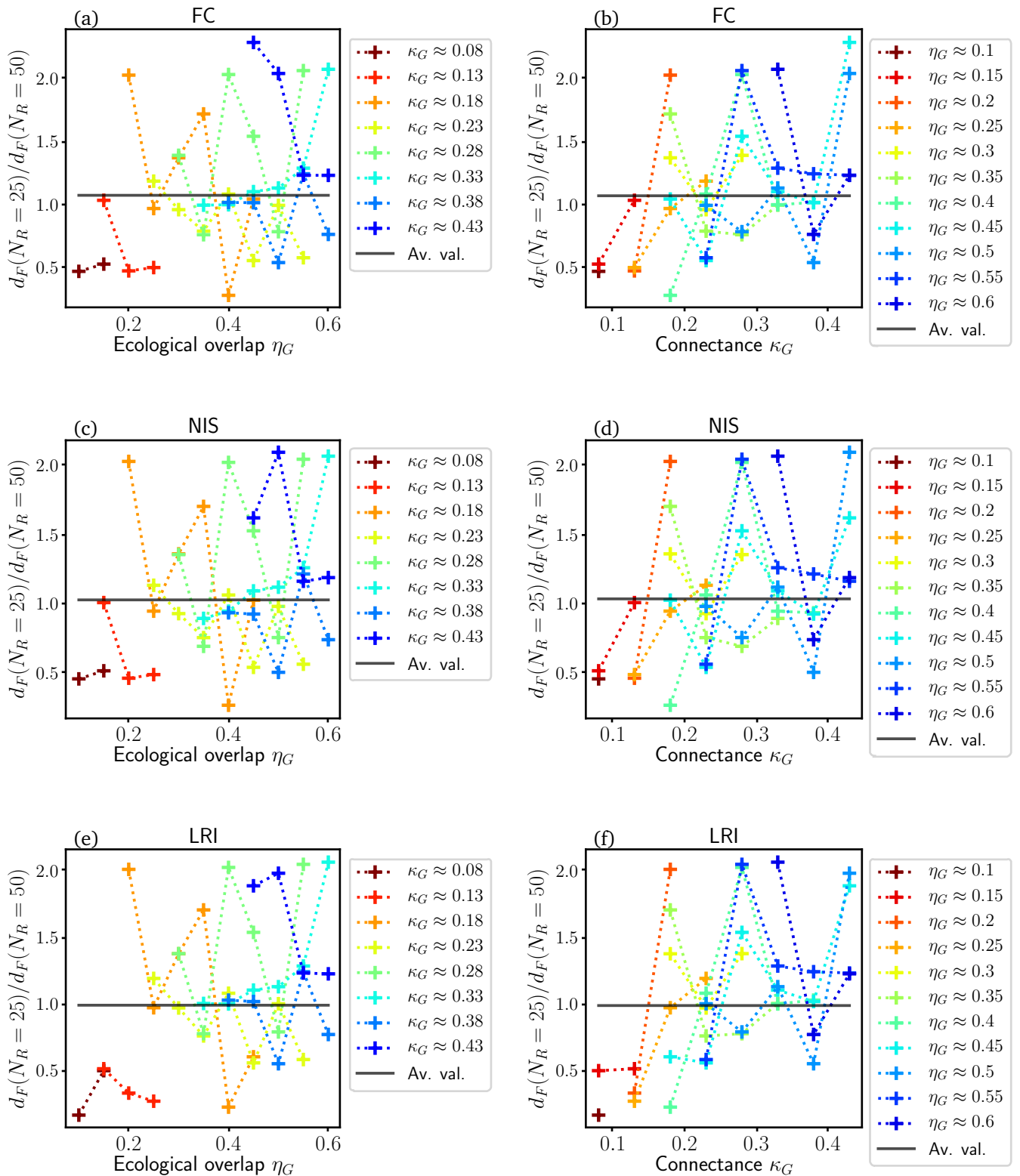


Figure 3.1.11: Common feasibility region  $\mathcal{F}_1^{SM}(\alpha_0)$  for  $N_R = 50$  and  $N_S = 25$ , to compare with 3.1.8 We considered different structures of the syntrophy matrix: (a) fully connected, (b) intraspecific syntrophy restricted and (c) LRI matrix. As the number of resources increases,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



we cannot say much from these figures. I think that perhaps our system has too many resources, all of them are supplied externally and at the needed rate, so perhaps this is why they do not add much

Figure 3.1.12: Ratio of the feasibility decay rates at  $N_R = 25$  and at  $N_R = 50$  as a function of the consumption matrix properties. A y-axis larger than 1 means  $d_F(N_R = 25)$  is larger than  $d_F(N_R = 50)$ , which means the system endures the addition of syntrophic interaction better at  $N_R = 50$ . We considered the four usual A scenarios (a)-(b) FC, (c)-(d) NIS, (e)-(f) LRI and (g)-(h) RS. Increasing the number of resources in the system does not allow microbial communities to be “more feasible” as syntrophy increases: on average  $d_F$  is unchanged by doubling the number of resources. A detailed how the consumption matrix properties, at least connectance and ecological overlap, or the A-scenario precisely modify the improvement is difficult to draw from this data. **TO DO: put also the one for random structure**

## 3.2 Dynamical stability

**TO DO:** write that even though feasibility does not imply dynamical stability globally does not mean that at certain points this is not true, say that large  $\lambda_1$  implies recovering from perturbations faster, mention somewhere that not fully feasible points not considered for dynamical stability

We discovered in the previous section that we expect microbial communities where a large syntrophic interaction is observed to have a large consumption rate  $\gamma_0$  and a small abundance of consumers at equilibrium  $S_0$ . It is time to work with feasible systems and investigate their dynamics. ~~We now spend our energy on studying the local dynamical stability of feasible microbial communities. In short, we want to know what happens to them~~ when their populations (of both resources and microbes) at equilibrium are perturbed. The question of the stability of ecosystems has always been and still remains of prime importance for ecologists and constitutes a whole scientific field. Many results have been derived, for linear systems [24], Lotka-Volterra systems [42, 43], mutualistic systems [19] or MacArthur's consumer-resource model [44]. An important trend, which we will check with our results, is ~~that~~ <sup>if</sup> highly connected ecological networks <sup>seem to be more robust</sup> to perturbations. **This was proven for structural stability, is it also true for dynamical stability?**

investigate system's behaviour

structurally stable

I don't think there is a general proof actually, if the network is random connectance may hinder dynamical stab.

In the next section we first discuss about the MCMC algorithm (Methods 2.2.4) designed to shape more dynamically stable systems. We then move on to study the dynamical stability of microbial communities by examining the shape of their fully dynamically stable regions, looking in what case feasibility implies dynamical stability and checking how the largest eigenvalue changes with the shape of the consumption-syntrophy network. Finally we will investigate what happens when the number of resources in the community is changed.

and see my comment in note 27

### 3.2.1 LRI regime – Outcome of the MCMC algorithm

In Methods section we explained how to find, for a given consumption matrix...

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:

fulfill conditions

- 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  $\mu$ , we expect the algorithm to yield  $A$ -matrices that minimize intraspecific syntrophy.
- The sum of the off-diagonal elements of  $|\alpha_0 AG - \gamma_0 R_0 G^T G|$  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  $(\mu, \nu)$ ,  $(AG)_{\mu\nu} = \sum_{i=1}^{N_S} A_{\mu i} G_{i\nu}$  is the number of species that at the same time consume resource  $\nu$  and release resource  $\mu$  and  $(G^T G)_{\mu\nu} = \sum_{i=1}^{N_S} G_{i\mu} G_{i\nu}$  is the number of con-

<sup>27</sup>This was demonstrated explicitly for mutualistic systems by Pascual-García and Bastolla in [19].

Not in general, only if the intraspecific competition was below a critical value, otherwise the connectance hinders structural stability and the nestedness is beneficial (difference between figure 2a and 2b in our paper)

I would write  $\alpha AG = \gamma R GG'$  because I think it is intuitive to see this in terms of syntrophy vs consumption. I realize now that what we discussed about using a unit vector  $\alpha u AG u' \sim \gamma R u GG' u'$  (excluding diagonals) and think about it in terms of the total flux of syntrophy vs the one of competition is not true due to the absolute value.

I think that the best intuition to me comes when  $AG \sim 0$  and  $GG' \sim 0$ .  $GG' = 0$  results when each species consumes only one (and different) resources. If they secrete only one result and there is no pair of species secreting the same, the sum of off diagonals is equal to  $Ns \cdot \alpha$ , if  $\alpha \ll Ns$  the sum is small. Let's think now that each species can consume only two resources. Then  $GG'$  again is minimized if there

### 3.2. DYNAMICAL STABILITY

### CHAPTER 3. RESULTS

are not two species consuming the same pair of resources. But then, for each species consuming one pair at rate  $\gamma_0 R$  there must be several species consuming one of the resources and secreting the other one  $\Rightarrow$  to maintain stability if I steal you I must compensate you

summers that eat both  $\nu$  and  $\mu$ . **TO DO: add something here, transition is a bit rough**

There are still things I do not understand. For instance, for each pair of resources the GG terms are counted twice (above and below the diagonal) and they are the same species consuming, while for the terms AG the same pair of resources must involve different species for consumption and secretion of

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  $(\mu, \nu)$ , the number of consumers that eat both  $\mu$  and  $\nu$  is proportional to the number of consumers species that eat  $\mu$  and release  $\nu$ . The proportionality constant, which is the same for all  $(\mu, \nu)$  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  $(\mu, i)$  if  $(i, \mu)$  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.

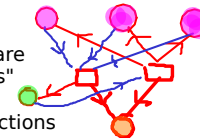
Figure 3.2.2 shows that indeed we obtain for a given  $G$  matrix a syntrophy matrix  $A$  such that the two requirements above are best satisfied. Note that the algorithm works better for matrices with a low connectance. **TO DO: explain why** It is worth noticing that this procedure produces highly 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**

The above reasoning indicates that to compensate one species being stolen several species releasing are needed so an algorithm willing to reach a LRI regime should consider the increase in the connectance of the syntrophy matrix, and low gamma connections would have a better outcome

### 3.2.2 Typical dynamical stability

"Typical"??

colors of the species are like "functional groups" meaning that they have the same connections



blue line = secretion  
red line = consumption  
"circles" = species  
"squares" = resources

We observed in Results 3.1.1 that, for all  $(G, A) \in S_{25}$ , at fixed  $\alpha_0$  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  $(\gamma_0, S_0)$  plane. Figure 3.2.4 shows that a similar although more complicated behaviour is observed in the case of the dynamical stability function  $\mathcal{D}_L((\gamma_0, S_0, \alpha_0), G, A)$ . On one hand, the  $(\gamma_0, S_0)$  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  $N_{\text{sys}}$  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  $N_{\text{sys}}$  would allow to reduce the relative uncertainty on  $\mathcal{D}_L$  and truly discover the origin of this interesting phenomenon.

I don't think the patchiness thing is very interesting given that we are talking about 98% or so of systems dynamically stable, with the colourmap is enough to show the phenomena and I think it comes from numerical effects indeed, we are also adding some noise so how knows

### 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<sup>28</sup> as

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

Intuitively,  $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$  corresponds to the set of all  $(\gamma_0, S_0)$  such that  $\mathcal{A}((\gamma_0, S_0, \alpha_0), G, A)$  is a feasible, locally dynamically stable parameters set with probability 1. Since we require  $\mathcal{A}((\gamma_0, S_0, \alpha_0), G, A)$  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 a naive approach, one could 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 3.2.5). However, as said before, because of the patchy 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. This means that contrarily to the case of feasibility, we must find another, more subtle, approach to study dynamical stability. We have to consider each consumption-network individually.

As we saw,  $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$  is geometrically more complex than  $\mathcal{F}_1^{G,A}(\alpha_0)$  (Figure 3.2.6) because of the *almost* fully dynamically stable points<sup>29</sup>. It may sometimes have holes, even without syntrophy, and sometimes not, even for matrices that are topologically very close. Compare for instance Fig 3.2.6a with Fig 3.2.6c, 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 3.2.6a) while the second can endure basically any feasible syntrophic interaction (Fig 3.2.6c). As a general trend, points with a larger  $\gamma_0$  and a smaller  $S_0$  remain dynamically stable (Fig 3.2.7). The question is then: why do other points lose their dynamical stability? Is it because they become unfeasible, or do they remain feasible but become dynamically unstable?

To answer that question, we quantify for each network  $(G, A) \in S_{25}$  the *probability of being dynamically stable when feasible*, denoted  $\text{Prob}(\mathcal{D}_L | \mathcal{F})(G, A, \alpha_0)$ . That quantity, formally defined in Appendix 5.2.8, has a straight forward interpretation: if for a certain  $(G, A)$ ,  $\text{Prob}(\mathcal{D}_L | \mathcal{F})(G, A, \alpha_0) = x$ , then a feasible  $(\gamma_0, S_0, \alpha_0)$  has on average a chance  $x$  to be dynamically stable<sup>30</sup>. Figure 3.2.8 shows that  $\text{Prob}(\mathcal{D}_L | \mathcal{F})(G, A, \alpha_0) < 1 \forall (G, A) \in S_{25}$  and  $\forall \alpha_0 \geq 0$ . That result has a very important biological consequence: for no consumption-syntrophy network **and at no syntrophy are all feasible systems dynamically stable**, i.e. there can always exist feasible but dynamically unstable microbial communities<sup>31</sup> **TO DO**

<sup>28</sup>A formal definition of the fully dynamically stable region  $\mathcal{D}_{L,1}^{G,A}$  is provided in Methods 2.2.1

<sup>29</sup>One may argue then that we should also consider the almost fully dynamically stable points in the analysis. That position is intellectually appealing but would require to either work in a completely different framework or choose a “stability threshold” which separates almost fully feasible points from the others. The arbitrariness of said threshold (should we take into account points above  $\mathcal{D}_L = 0.99$ ? Or above  $\mathcal{D}_L = 0.98$ ? What about 0.995?) would make such a point of view hard to hold.

<sup>30</sup>More precisely, there is a chance  $x$  that a parameters set  $\mathcal{A}((\gamma_0, S_0, \alpha_0), G, A)$ , which we know is feasible, is also dynamically stable.

<sup>31</sup>This does not mean they will survive, of course, since they can vanish at any small perturbation. However it means that there is no physical law that hinders their existence: they can exist, and could appear e.g. through mutation processes.

54 As far as I understand is what you do in figure 3.2.8, there is no need of establishing a threshold just showing what you obtain. To me it is more arbitrary to open complicated questions about patchiness when it is possibly a consequence of the noise and/or numerical errors

Syntrophic interaction in microbial communities

I find this analysis misleading, because you are considering F1, what if the system you are saying loses dynamical stab. has F0.99? I would not say that there are so dramatic differences.

Why you are not showing the same "problematic nets" you were showing above?

review wording



: if time left, do a fit of the slope but not sure if that is very relevant or what it can tell (rate at which you lose dynamical stability?)-> would show that LRI improves stuff at low connectance.

**TO DO : talk about the fact that LRI does not improve anything**

Similarly to what was done for feasibility in Results 3.1.2 we can measure the volume of dynamical stability **TO DO: check this is the right name** (see Appendix 5.2.6)  $\text{Vol}(\mathcal{D}_{L,1}^{G,A}(\alpha_0))$  of each consumption-syntrophy network  $(G, A)$ .  $\text{Vol}(\mathcal{D}_{L,1}^{G,A}(\alpha_0))$  tells us what proportion of the unit square is occupied by dynamically stable  $(\gamma_0, S_0)$  points and is therefore an indicator of how well communities can sustain an increase of average syntrophic strength  $\alpha_0$ . Figure 3.2.9 shows that the dynamically stable volume typically decays in an exponential-like fashion. We inspire ourselves from the work of Results 3.1.2 to quantify that decay, and define the *dynamical stability decay rate*  $d_D(G, A)$  of a consumption-syntrophy network  $(G, A)$ . It is obtained numerically by finding through a non-linear regression the coefficients  $c_1, c_2, d_D \in \mathbb{R}^+$  that satisfy best the relation:

$$\text{Vol}(\mathcal{D}_{L,1}^{G,A}(\alpha_0)) \approx c_1 \exp(-d_D \alpha_0) - c_2. \quad (3.7)$$

Much like for the corresponding discussion in the case of feasibility,  $d_D(G, A)$  is an indicator of how well a consumption-syntrophy network  $(G, A)$  can sustain syntrophy while remaining dynamically stable: a smaller  $d_D(G, A)$  indicates a more robust network (it will take a larger syntrophy to find unstable points), a larger  $d_D(G, A)$  shows the network is weak against an increase in syntrophic interaction. Figure 3.2.10 shows how  $d_D(G, A)$  changes as a function of the characteristics of the consumption matrix  $G$ ,  $\forall G \in G_{25}$ . A clear trend may be seen: in order to avoid losing dynamical stability when syntrophy is increased, a microbial community should either increase the number of average resources eaten by each consumer (i.e. increase the connectance of the consumption matrix) or decrease its ecological overlap, which means that consumers should stop eating from the same resources (an intuition on why this makes sense biologically is provided in the next section **TO DO: rewrite something here better**). Figure 3.2.11 shows how  $d_D(G, A)$  changes for each matrix as different syntrophy regimes are considered. The following comments can be made<sup>32</sup>:

- The NIS scenario does not significantly decrease the dynamical stability decay rate. It consistently decreases it by  $\sim 5 - 15\%$  **TO DO: check errors**, except in the case  $\kappa_G = 0.18, \eta_G = 0.15$ , where  $d_D(G, A)$  is decreased by  $\sim 30\%$ .
- The LRI scenario greatly improves ( $> \sim 50\%$ ) the dynamical stability for consumption matrices with a very low connectance. On average, the lower the connectance, the greater the improvement. However, for larger connectances, it does not change  $d_D(G, A)$  in any way. *the effect saturates maybe?*
- The RS scenario offers the best improvement: any consumption matrix we considered got a considerably better result (at least  $> \sim 20\%$ ). At fixed connectance, the

Again, I do not think this is small

Following the same line of reasoning

<sup>32</sup>In the following, we use terms like “is better”, “outperforms” or “improves the dynamical stability” as a way of saying “lowers the dynamical stability decay rate compared to the FC case”.

improvement is the same for every ecological overlap considered and the lower the connectance the better the improvement. And we possibly understood why above

Overall, the LRI and RS scenarios offer a **more significant improvement** the lower the connectance of the consumption matrix. Because both have a syntrophy matrix whose connectance is equal to the one of the consumption matrix, this suggests that microbial communities which have very few syntrophic interactions (*i.e.*  $A$  has a low connectance) can remain dynamically stable with a larger average syntrophic interaction than others. Because the random structure (RS) scenario is, for a given matrix, a better improvement than LRI, we think that the syntrophy matrix should also have a lower syntrophic overlap, *i.e.* a low nestedness<sup>33</sup> **TO DO: write why that makes sense, you depend less on the others who are changed with the perturbation**

### 3.2.4 Largest eigenvalue of the jacobian

Equation (2.133) from Methods 2.2.5 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 we should<sup>34</sup>

- Decrease  $N_S$  or  $\alpha_0$ .
- If  $\alpha_0 - \gamma_0 R_0 < 0$ , increase  $N_R$ ,  $\sigma_0$  and  $\gamma_0$ .
- Be careful in how are you teaching? :-) you handle  $S_0$ : Appendix 5.1.2 shows that at very low syntrophy, we should increase  $S_0$  but decrease it when  $\alpha_0$  becomes **very significant**. Since in practice the observed feasible  $\alpha_0$  are fairly low, we should increase  $S_0$  as much as possible. reserved word for statistics

Combining these considerations with the feasibility conditions Eq. (2.28) we expect that – for all other metaparameters fixed – systems get more and more locally dynamically stable as  $\gamma_0$  is increased and  $S_0$  has its largest feasible value. In short, points at the upper border of  $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$  should have a lower and lower  $\text{Re}(\lambda_1)$  as  $\gamma_0$  increases. Figure 3.2.12 shows that **indeed** that trend is ~~indeed~~ observed. This tells us that if we keep the consumption flux  $N_S \gamma_0 S_0$  constant, increasing  $\gamma_0$  (and hence decreasing  $S_0$ ) will give rise to more stable systems. Notice that contrarily to the prediction made above, increasing  $\alpha_0$  does not decrease stability but increases the maximal  $|\text{Re}(\lambda_1)|$  observed as shows Fig. 3.2.13a. That is coupled with the already discussed shrinkage of the fully locally dynamically stable volume seen on Fig. 3.2.13b. 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. **TO DO : need to elaborate more?** I should see what comes out from structural stability first

<sup>33</sup>Indeed, the LRI regime has a significantly more nested syntrophy matrix compared to RS (Fig 3.2.3). At equal connectance,  $A$  with the lower nestedness provides a better improvement.

<sup>34</sup>We do not focus on how  $R_0$  and  $l_0$  should be changed because they are always equal to 1 for our study of feasibility and local dynamical stability.



### 3.2.5 The influence of the matrix dimension

As said above, because of Eq. (2.133), we expect dynamical stability to improve when the number of resources is increased and the number of consumers is kept fixed. It is therefore worth briefly studying what happens when the number of resources is doubled  $N_R = 25 \rightarrow N_R = 50$  and every other metaparameter, as well as the number of consumers, keeps the same value as before.

Figure 3.2.14 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  $\alpha_0$ . Interestingly, adding more resources seems to take away the patchiness of  $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$ , which in consequence drastically changes the common locally dynamically stable region. Indeed  $\mathcal{D}_{L,1}^{S_{50}}(\alpha_0)$  is smoother than  $\mathcal{D}_{L,1}^{S_{25}}(\alpha_0)$  and can sustain a non-zero syntrophy (see the striking difference between Figure 3.2.5 and Figure 3.2.16). Recall that for  $N_R = 25$ , the critical common syntrophy was between 0 and  $1.3 \times 10^{-3}$ . It is greatly improved for  $N_R = 50$ : between  $3.9 \times 10^{-3}$  and  $5.2 \times 10^{-3}$ .

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 3.2.15b with Fig 3.2.13b). 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 3.2.15a and Fig 3.2.13a). 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, **Is there anything more to add? I have some plots but I am not sure if they are the most relevant TO DO: check that last paragraph and put discussion about decay rate comparison**

In figure 3.2.5 it is not clear which is the topology, do not see that different than 3.2.6, it should actually be quantified the volume via the sum of the probabilities rather than D1. Then you compare the rate of decay between 25 and 50 and are the same, but you claim that the volume is different...

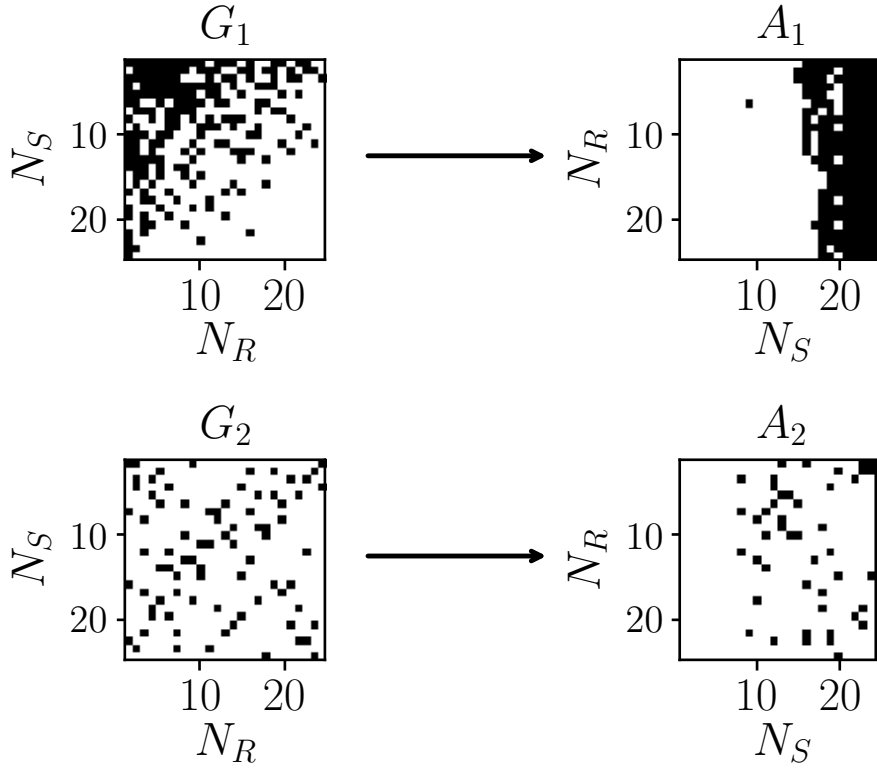
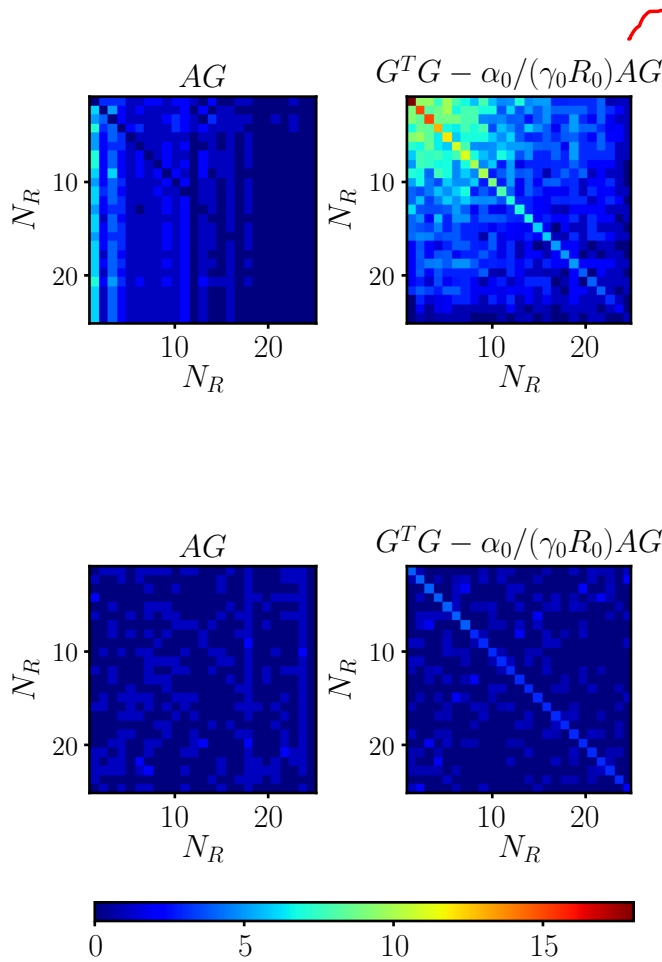


Figure 3.2.1: Typical shape of the consumption  $G_i$  and syntrophy  $A_i$  matrices. The white cells symbolize a zero matrix element and the black cells, a one.  $A_i$  here is the outcome of the LRI MC algorithm described in Methods [2.2.4](#). The first row has a consumption matrix with  $\eta_1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



To me this figure is still confusing, since only the off diagonal elements are relevant here why don't you make them zero and say it in the caption? I do not see that values outside the diagonal are zero, are as high as ~11-12

Figure 3.2.2: Plotting of  $AG$  and  $G^T G - \alpha_0 / (\gamma_0 R_0) AG$ . 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,  $G^T G - \alpha_0 / (\gamma_0 R_0) AG \approx 0$ . Both relations are better satisfied for consumption (and hence syntrophy) matrices with a low connectance.

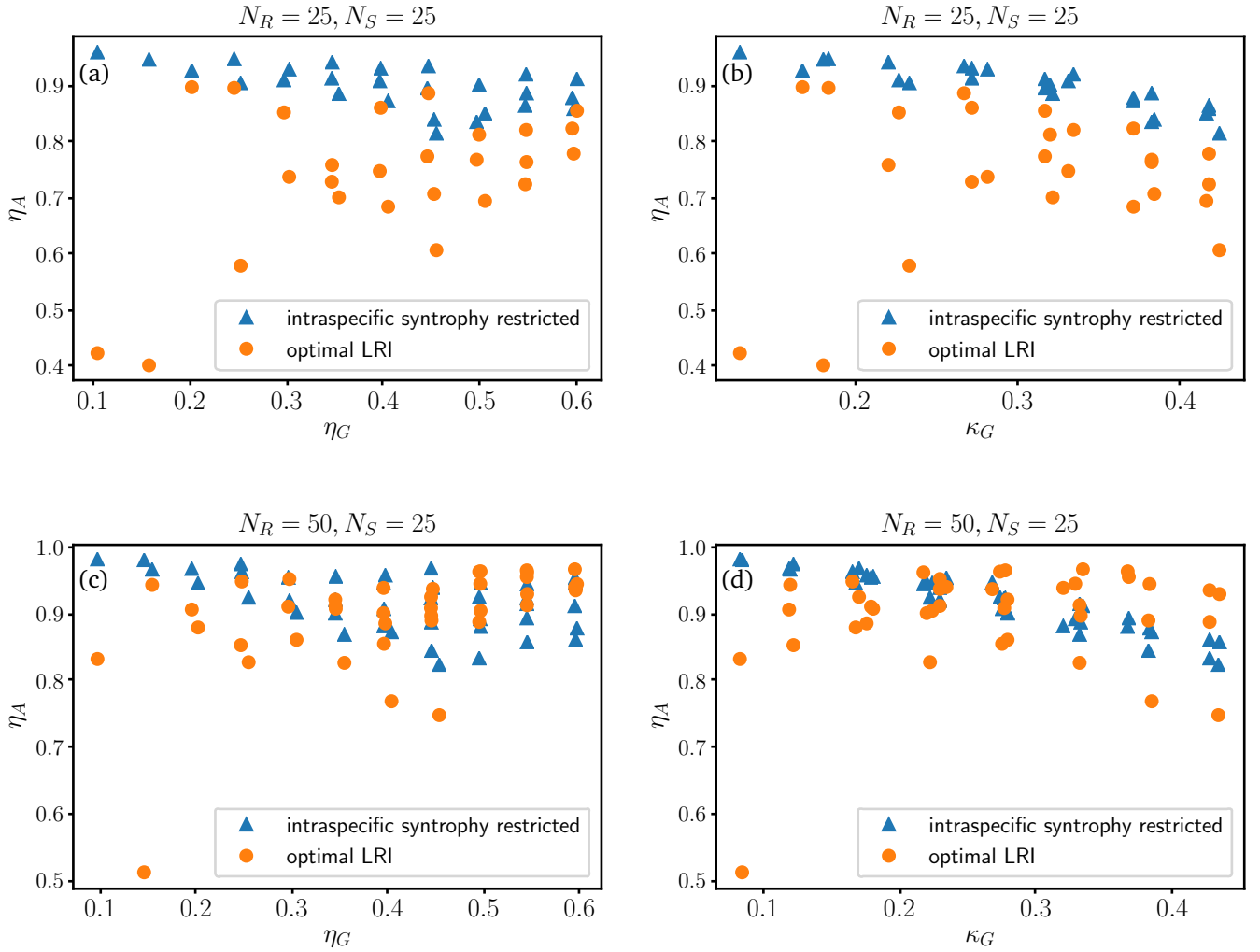


Figure 3.2.3: Properties of the syntrophy matrix against the consumption matrix. (a)-(c) Ecological overlap of  $A$  as a function of the ecological overlap of  $G$  for  $N_S = 25$  and  $N_R = 25$  (a) or  $N_R = 50$  (c). (b)-(d) Ecological overlap of  $A$  as a function of the connectance of  $G$  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  $\eta_G$  or  $\kappa_G$  increase, the two results will ~~without surprise~~ give matrices with similar properties. **Explain this?**

No, explain in the text if needed

Figure a is redundant --> remove

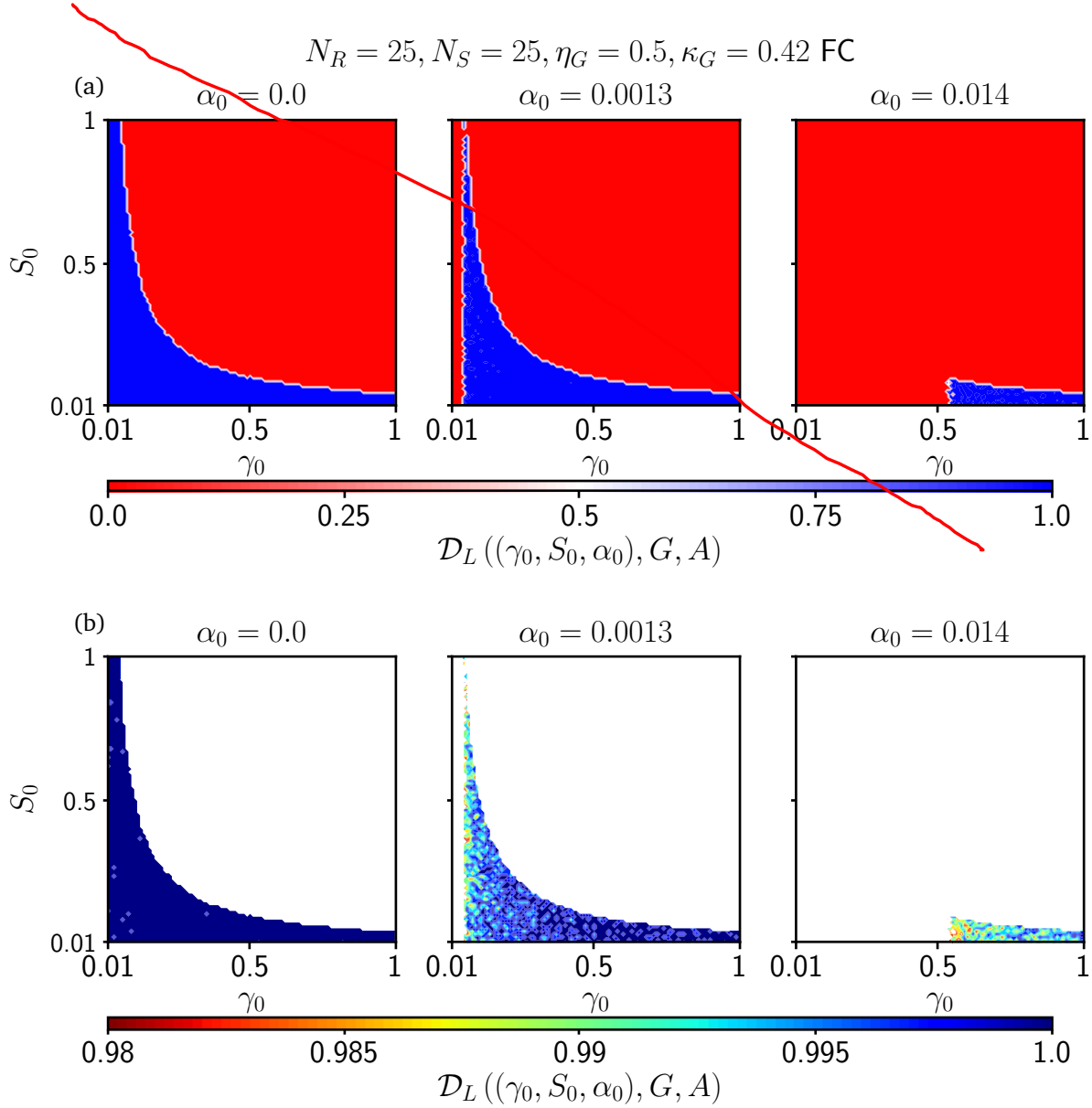


Figure 3.2.4: Typical color plot local dynamical stability metaparameters function  $\mathcal{D}_L$ . for microbial communities with  $\eta_G = 0.5$  and  $\kappa_G = 0.42$ . The color bar indicates the value of  $\mathcal{D}_L((\gamma_0, S_0, \alpha_0), G, A)$  with  $A$  fully connected. The plane 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.

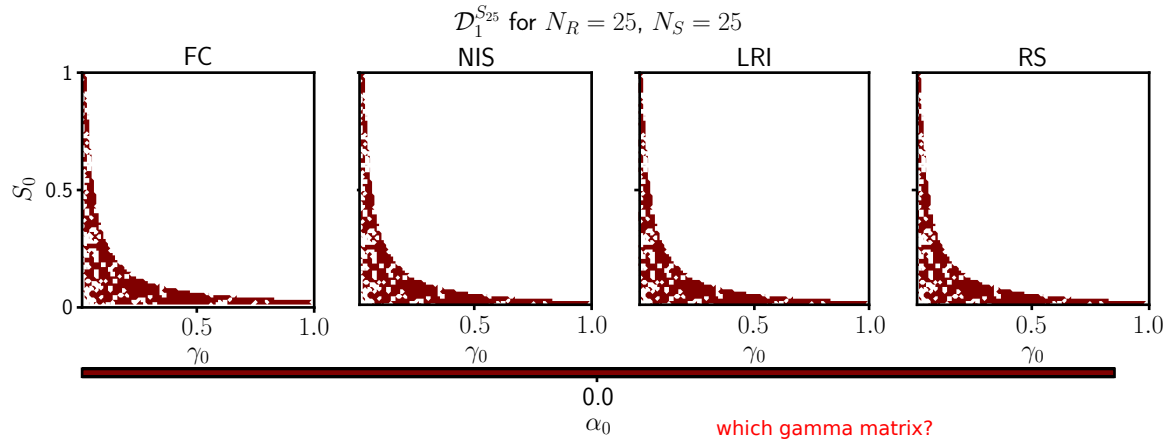


Figure 3.2.5: Common full local dynamical stability volume for different  $A$  structures: (a) fully connected, (b) no intraspecific syntrophy and (c) LRI algorithm. The points coloured in dark red give rise to locally dynamically stable systems with probability 1 for all the matrices considered. Very 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.

I think that instead of this figure, here should be one with the probability of having a dynamically stable system for three different values of alpha for example. It will have more information and it good be a better introduction for the next figure and for the comparison with  $N=50$

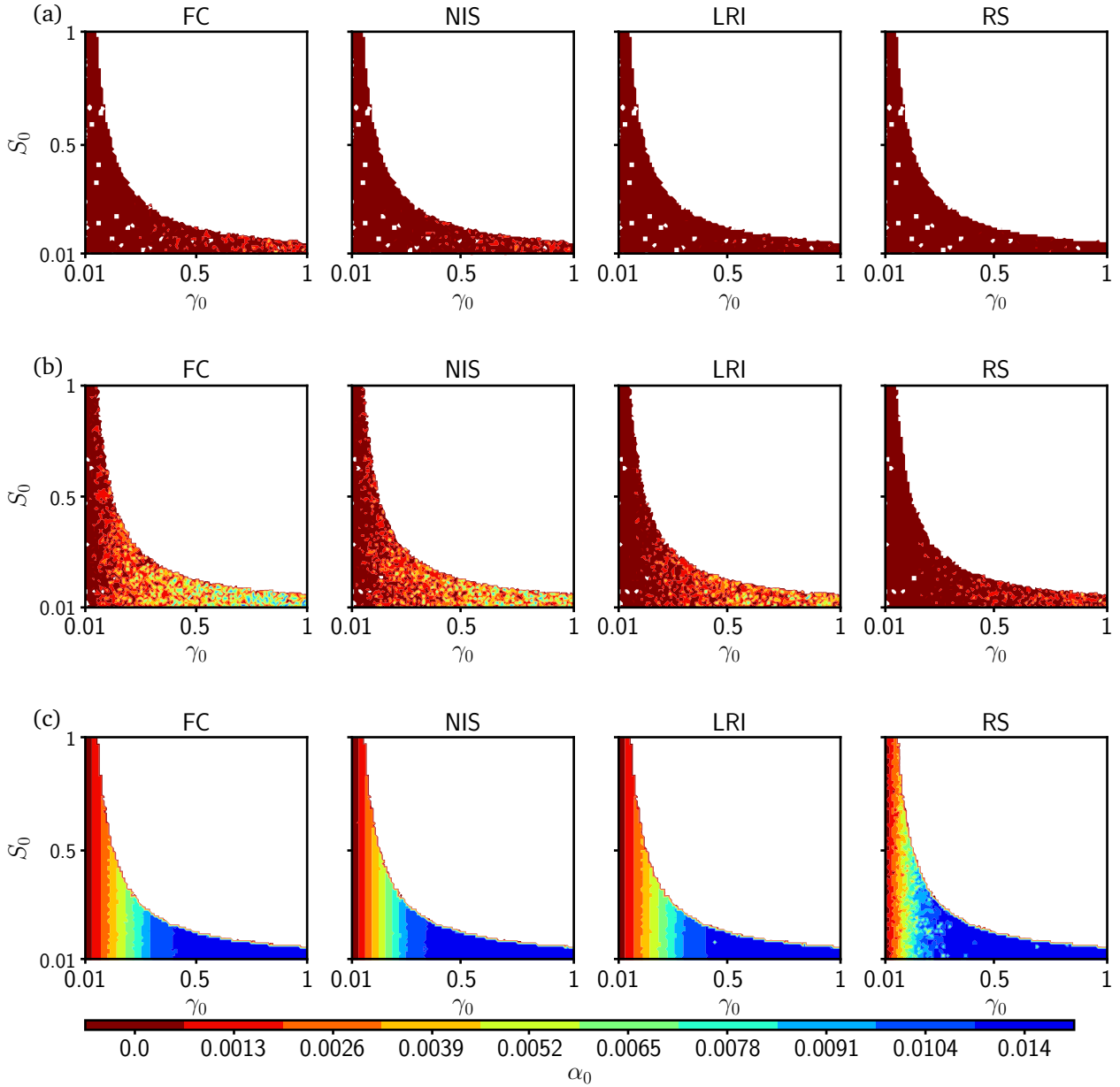
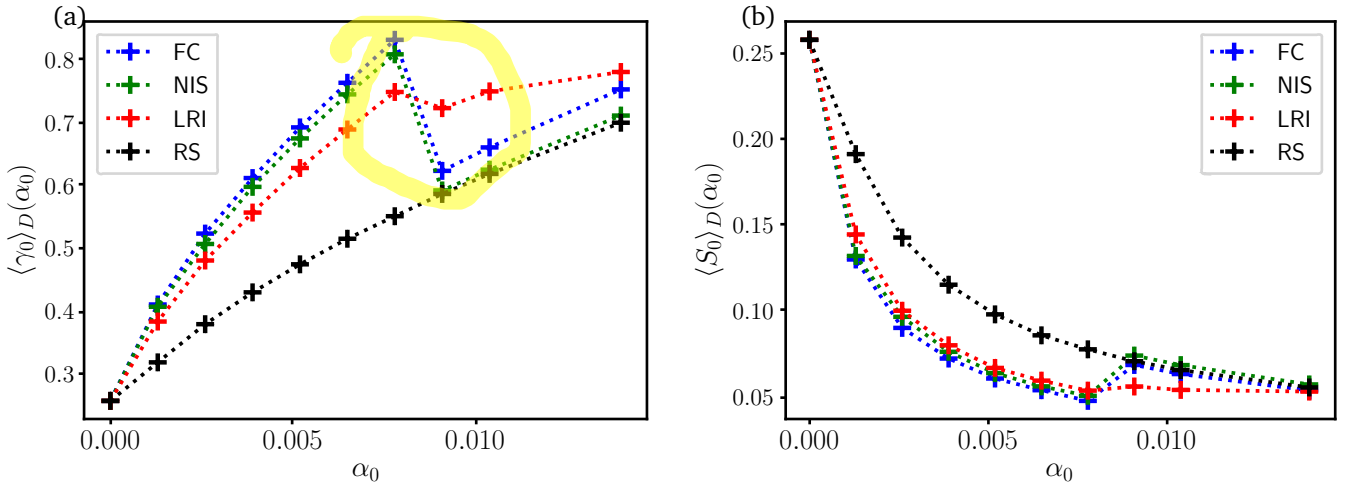


Figure 3.2.6: Locally fully dynamically stable region  $\mathcal{D}_{L,1}^{G,A}$  as a function of syntrophy for different matrices  $G$ . The white zone corresponds to points that are never fully locally dynamically stable. The colour of a given point tells until which syntrophy that point is fully locally dynamically stable, e.g. a green point is fully locally dynamically stable for  $0 \leq \alpha_0 \leq 6.5 \times 10^{-3}$ . 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.



overloaded notation, when you plot  $y$  vs  $x$  do you write  $y(x)$  everywhere?

Figure 3.2.7: (a) Average consumption rate  $\langle \gamma_0 \rangle_D(\alpha_0)$  and (b) average resource abundance equilibrium  $\langle S_0 \rangle_D(\alpha_0)$  as a function of syntrophy (a formal definition of  $(\langle \gamma_0 \rangle_D(\alpha_0), \langle S_0 \rangle_D(\alpha_0))$  is provided in Appendix 5.2.7). Intuitively,  $(\langle \gamma_0 \rangle_D(\alpha_0), \langle S_0 \rangle_D(\alpha_0))$  represents the center  $(\gamma_0, S_0)$  point of  $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$ , averaged over all matrices  $(G, A) \in S_{25}$ , or as we call it the “center of dynamical stability”. As syntrophy increases, only points with a large consumption rate and a small resource abundance at equilibrium remain dynamically stable. The sudden drop of  $\langle \gamma_0 \rangle_D$  (and rise of  $\langle S_0 \rangle_D$ ) at  $\alpha_0 = 9.1 \times 10^{-3}$  is a finite-size effect. Indeed we only monitor points in the unit square  $[0, 1]^2$ , such that the  $(G, A)$  for which the center of  $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$  leave that zone at  $\alpha_0 = 9.1 \times 10^{-3}$  do not contribute to  $(\langle \gamma_0 \rangle_D, \langle S_0 \rangle_D)$  anymore. Only the  $(G, A)$  which previously had a lower, resp. higher, contribution to  $\langle \gamma_0 \rangle_D$ , resp.  $\langle S_0 \rangle_D$ , are taken into account, which results in that strange behaviour. The fact that  $\langle \gamma_0 \rangle_D$  continues to increase (and  $\langle S_0 \rangle_D$  to decrease) after that point show that this reasoning is right.

Move to text

This is the predicted LRI behaviour right?

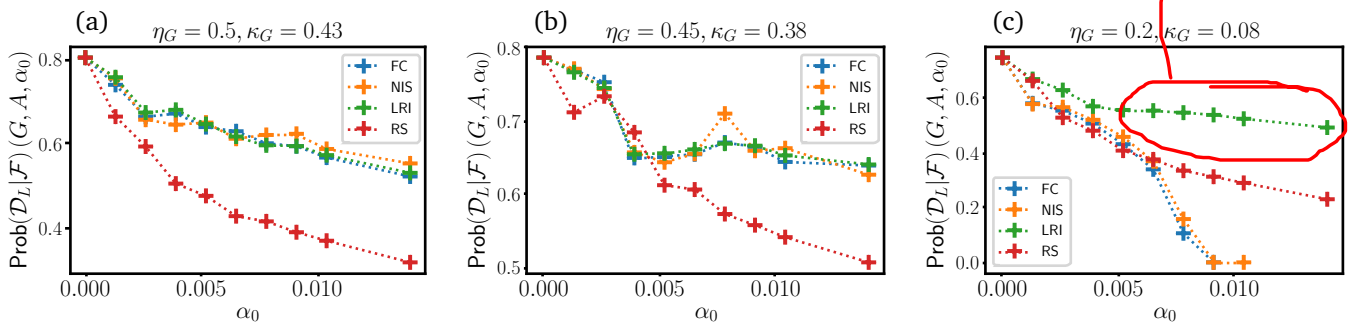


Figure 3.2.8: Plot of the probability that a microbial community is dynamically stable ~~under the assumption that it is feasible~~ for different consumption-syntrophy networks  $(G, A)$ . The different lines on the same subplot show the four different  $A$ -scenarios. The results differ with the consumption matrix considered: (a)  $\eta_G = 0.5, \kappa_G = 0.43$ , (b)  $\eta_G = 0.45, \kappa_G = 0.38$  and (c)  $\eta_G = 0.2, \kappa_G = 0.08$ .

Since you presented three problematic matrices above that were very similar, we want to see them also here



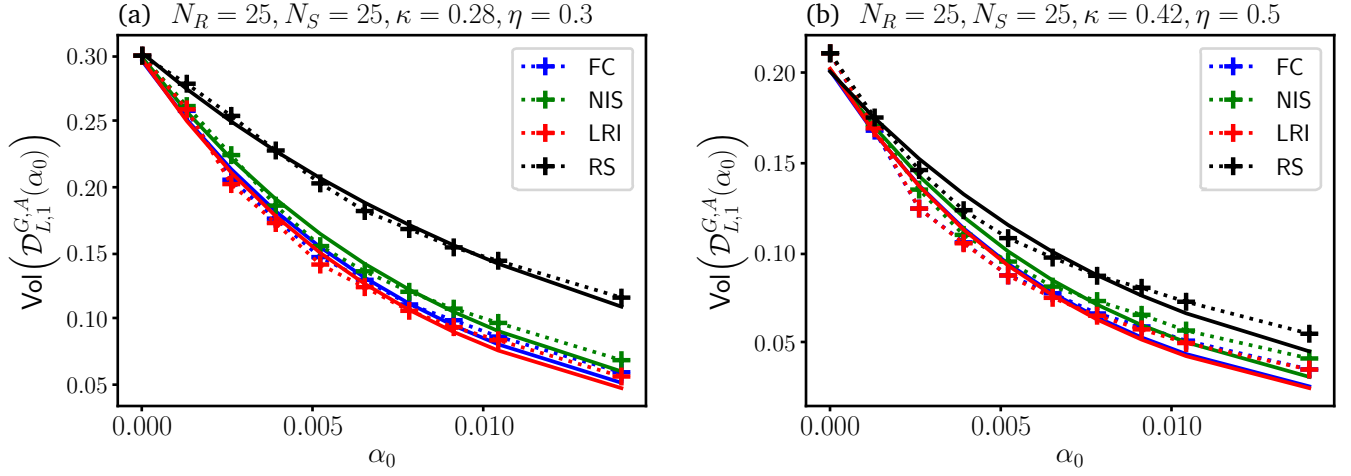


Figure 3.2.9: Evolution of the dynamically stable volume  $\text{Vol}(\mathcal{D}_{L,1}^{G,A}(\alpha_0))$  (see Appendix 5.2.6) with  $\alpha_0$  for different  $(G, A) \in S_{25}$ . (a)  $\eta_G = 0.3$  and  $\kappa_G = 0.28$ , (b)  $\eta_G = 0.5$  and  $\kappa_G = 0.43$ . The cross-shaped points indicate the data as measured numerically while the solid lines are the corresponding exponential fit (see main text). The four colors indicate the four  $A$ -scenarios considered. Independent of the  $(G, A)$  network, fewer points are dynamically stable as syntrophy increases.

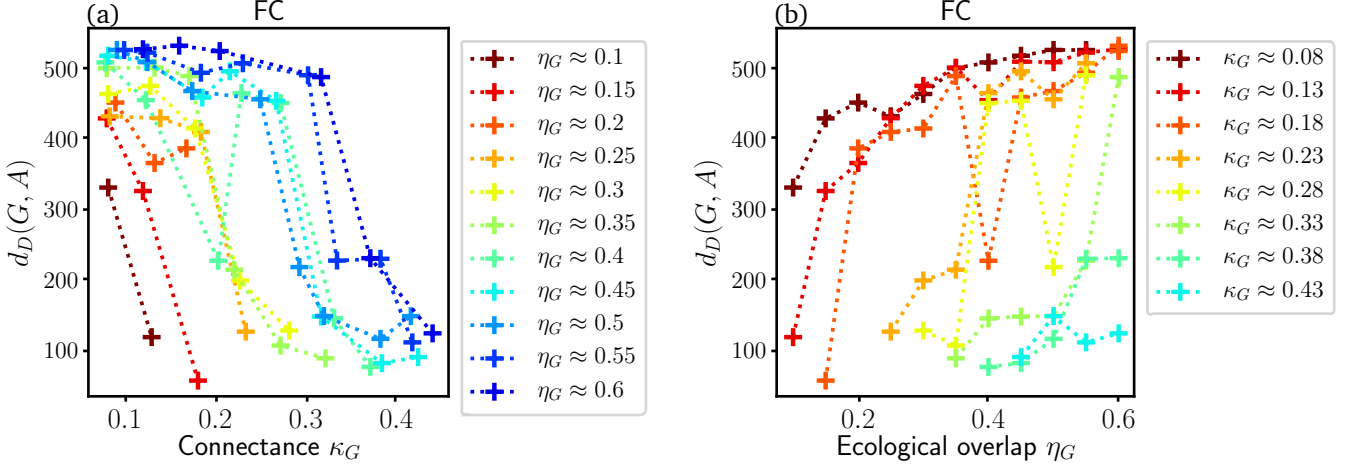


Figure 3.2.10: Dynamical stability decay rate  $d_D(G, A)$  for the  $A$ -matrix fully connected (FC scenario) and every  $G \in G_{25}$ , (a) as a function of connectance  $\kappa_G$  for fixed ecological overlap  $\eta_G$  and (b) as a function of ecological overlap for fixed connectance. The trend confirms the previous observations: at fixed ecological overlap, a microbial community with a more connected consumption matrix will sustain a larger syntrophy (i.e. have a smaller decay rate) and at fixed connectance, systems with a small ecological overlap remain dynamically stable as syntrophy grows.

All these figures are super similar to those of feasibility, did you try to plot one measure against the other? dF vs dD for instance

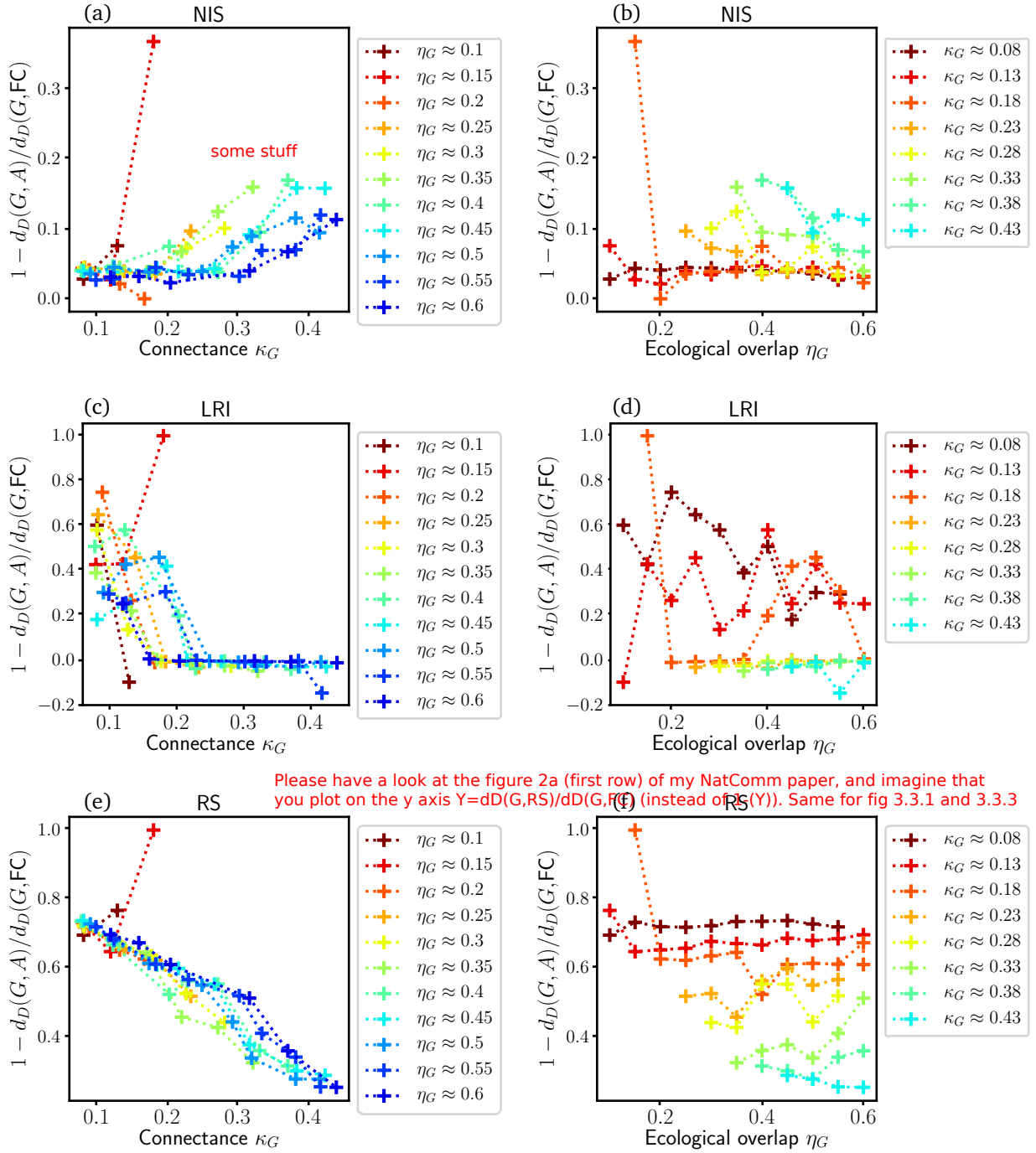


Figure 3.2.11: Relative deviation away from the FC case dynamical stability decay rate as a function of the characteristics of the consumption matrix  $G$ . A positive  $y$ -coordinate means that the considered scenario provides a  $d_D(G, A)$  smaller than the FC case and is in consequence a sign that the network can better withstand an increase in syntrophy while remaining dynamically stable. We considered the three usual scenarios for the  $A$ -matrix: (a)-(b) NIS, (c)-(d) LRI and (e)-(f) RS. **TO DO: compute errors on this graph**

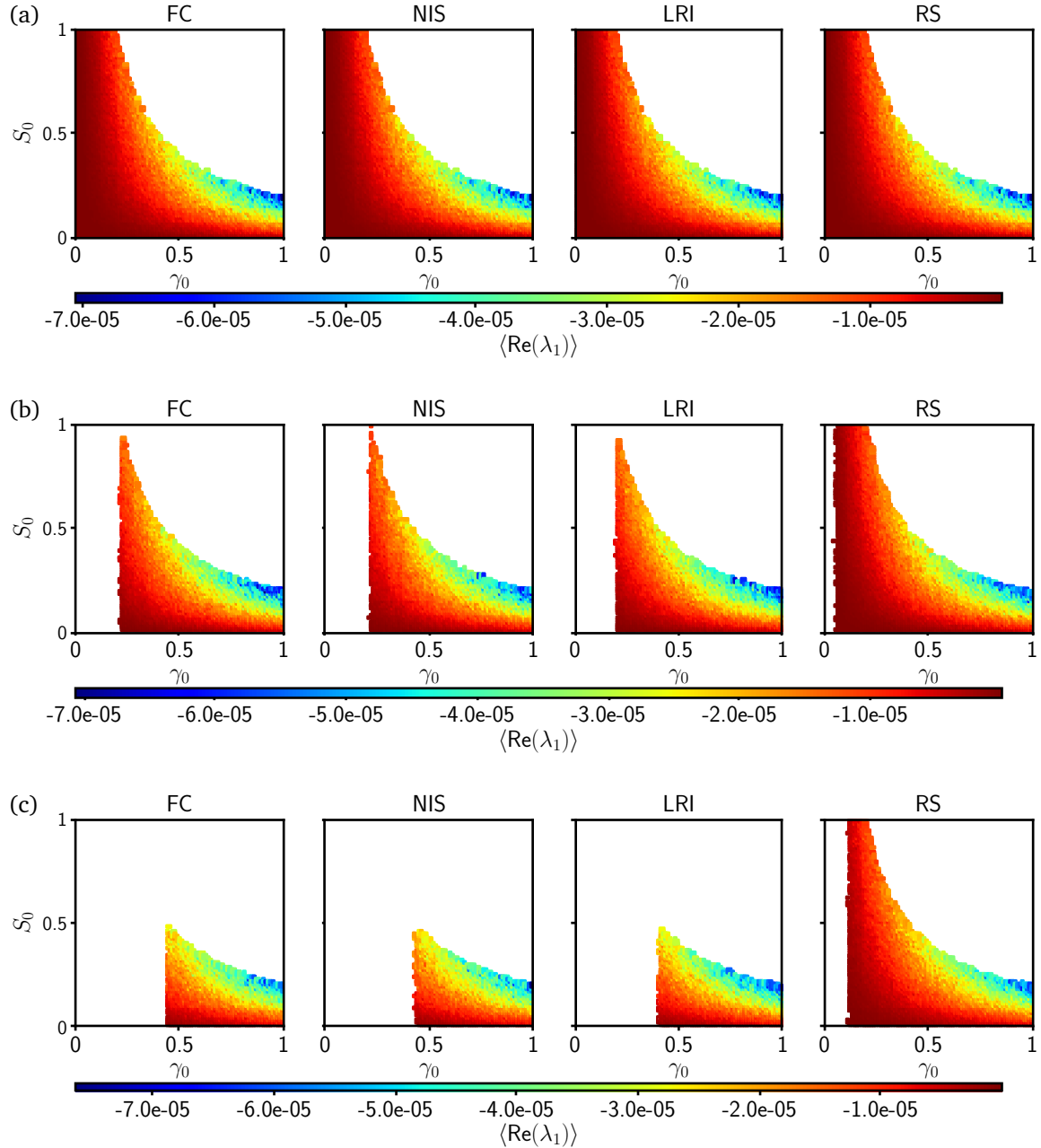


Figure 3.2.12: Largest real eigenvalue  $\text{Re}(\lambda_1)$  averaged over 300 **TO DO : check number** realisations for each  $(\gamma_0, S_0)$  points for the consumption matrix  $G$  with consumers overlap  $\eta_G = 0.1$  and connectance  $\kappa_G = 0.13$ . The white points correspond to not 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 different columns correspond to the different  $A$  scenarios considered, which change the shape of the dynamically stable region as seen in the main text. As expected, the boundary points close to the  $\gamma_0 \sim S_0^{-1}$  curve are the most stable in every situation. Points with a large  $\gamma_0$  are the most stable of the unit square.

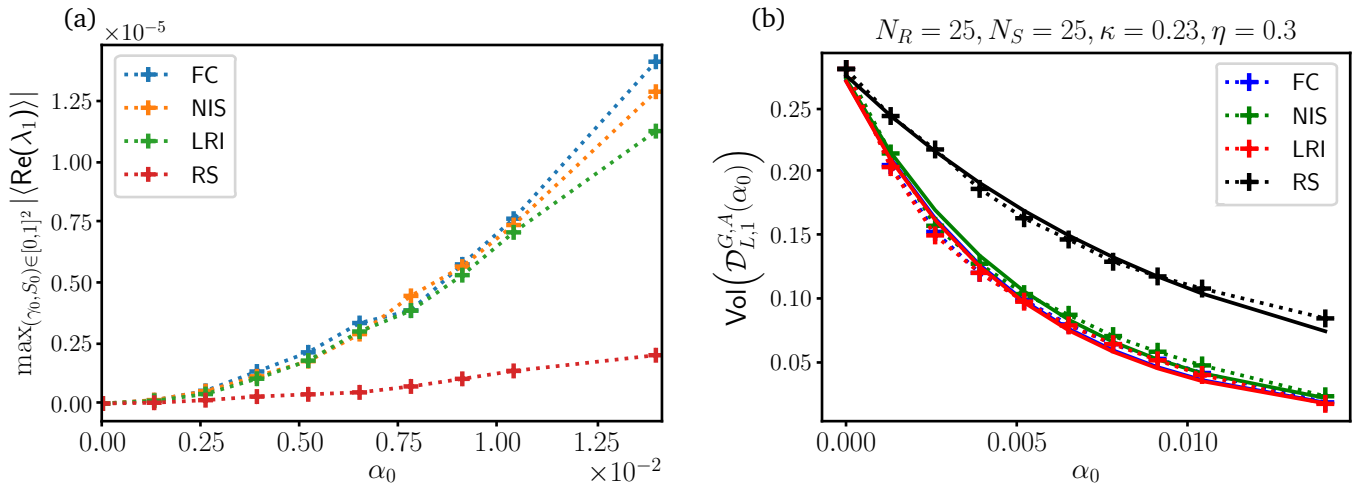


Figure 3.2.13: For a consumption matrix  $G$  with  $\eta_G = 0.3$  and  $\kappa_G = 0.23$ . (a) Evolution of the maximal  $|\langle \text{Re}(\lambda_1) \rangle|$  observed in the  $(\gamma_0, S_0) \in [0, 1]^2$  region. The maximal eigenvalue increases in magnitude, making the system more dynamically stable, as syntrophy increases. That trend is true for all matrices we considered. (b) Volume of  $\mathcal{D}_{L,1}^G(\alpha_0)$ . As syntrophy increases, fewer and fewer points become fully dynamically stable. For both figures, the different lines show the different structure of the syntrophy matrix that we considered. **TO DO: replace this with average version**

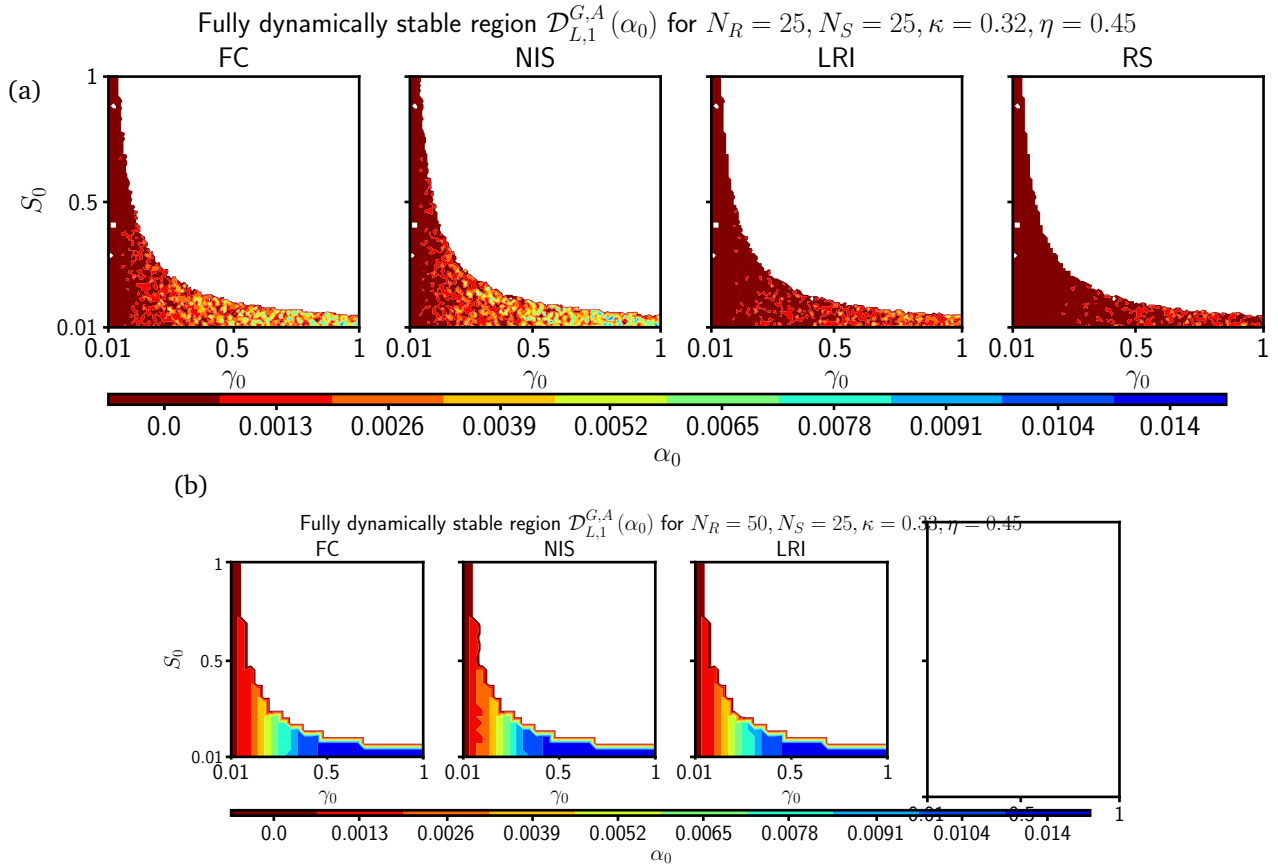


Figure 3.2.14: Fully dynamically stable region  $\mathcal{D}_{L,1}^{G,A}$  with the three different structures of  $A$  considered: fully connected (left), no intraspecific syntrophy (middle) and LRI matrix (right). The two matrices have the same ecological overlap and connectance, only the number of 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  $\eta_G = 0.45$  (by lack of time, a fewer resolution on the unit square had to be taken). At  $N_R = 50$ , the fully dynamically stable region seems less fractured. 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.**

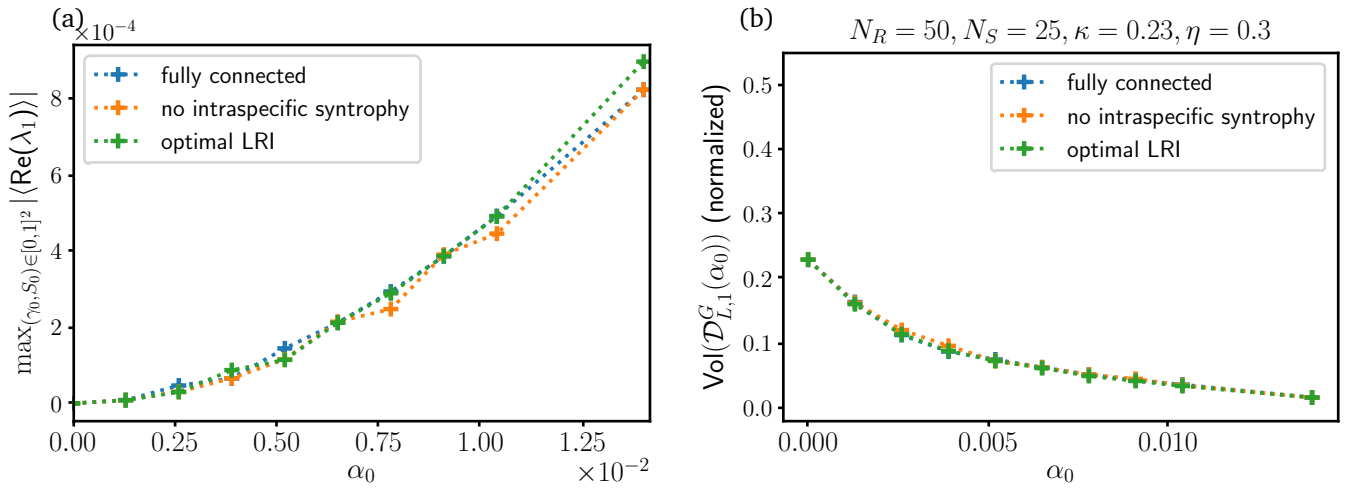


Figure 3.2.15: To be compared with Fig. 3.2.13. The consumption matrix  $G$  considered here has  $\eta_G = 0.3$  and  $\kappa_G = 0.23$ . (a) Maximal average  $|\text{Re}(\lambda_1)|$  observed in the unit square. (b) Percentage of the unit square occupied by the fully dynamically stable region of  $G$  as a function of syntrophy. The matrix considered has almost equal properties to the one in Fig. 3.2.13 with the only difference that  $N_R = 50$  here. Even though the size of  $\mathcal{D}_{L,1}^{G,A}$  is smaller, the eigenvalues are larger in magnitude.

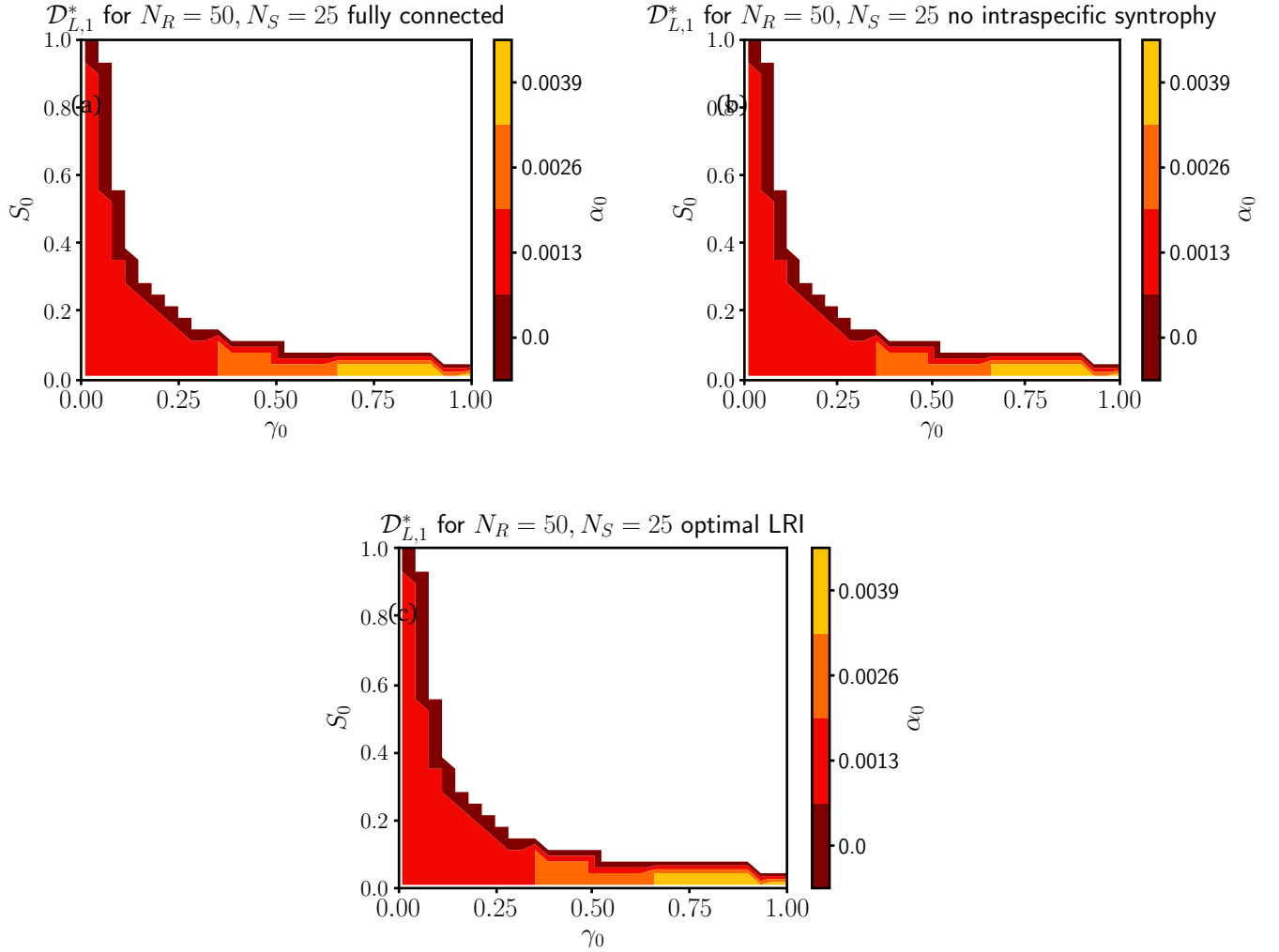


Figure 3.2.16: Common fully dynamically stable volume. It is larger with a larger number of resources -> even though individually it is not always better; it is better for the worse matrices (since the common volume can handle more syntrophy)

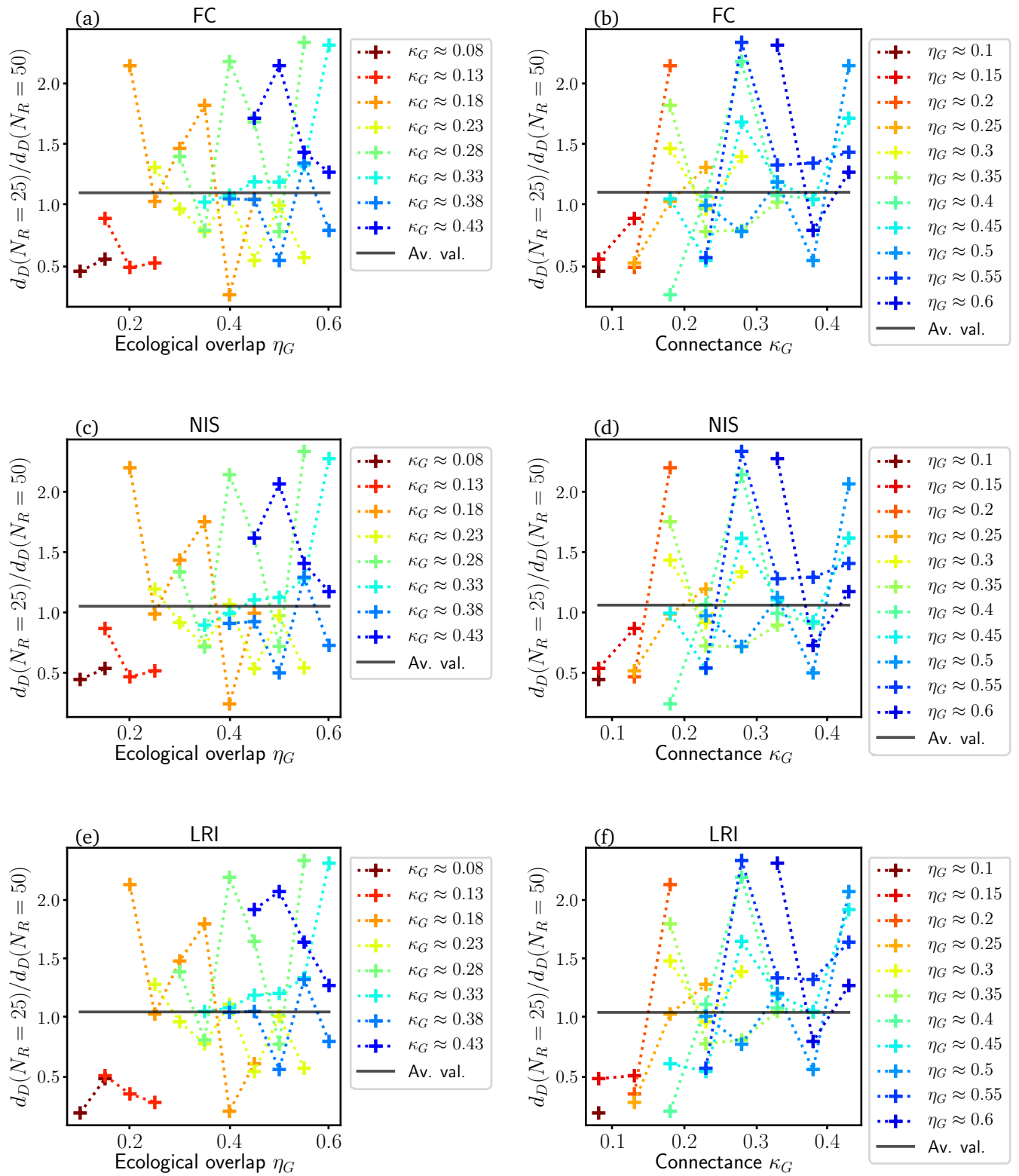


Figure 3.2.17: Ratio of the dynamical stability decay rates at  $N_R = 25$  and at  $N_R = 50$  as a function of the consumption matrix properties. A y-axis larger than 1 means  $d_D(N_R = 25)$  is larger than  $d_D(N_R = 50)$ , which means the system remains “more” dynamically stable as syntrophic interaction is added at  $N_R = 50$  compared to  $N_R = 25$ . We considered the four usual A scenarios (a)-(b) FC, (c)-(d) NIS, (e)-(f) LRI and (g)-(h) RS. On average increasing the number of resources in the system does not particularly allow microbial communities to be slightly “more dynamically stable” as syntrophy increases since the average ratio between the dynamical stability decay rates is very close to 1. A detailed on how the consumption matrix properties, at least connectance and ecological overlap, or the A-scenario precisely modify the improvement is difficult to draw from this data. **TO DO: put also the**



### 3.3 Structural stability

**TO DO : say NR=50 less patchy -> easier to work with**

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 [45, 46], 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<sup>35</sup>.

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  $\alpha_0$  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 5.2.9 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  $\sim 1$  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  $\alpha_0$  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  $(\gamma_0, S_0)$  such that  $(\gamma_0, S_0, \alpha_0) \in \mathcal{D}_{L,1}^{G,A}(\alpha_0)$  for all values of  $\alpha_0$  we want to investigate. The idea is of course to take  $\alpha_0$  as large as possible. Figure 3.2.5 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  $1.3 \times 10^{-3}$ , which is only a tenth of the largest common syntrophy  $\sim 0.01$  (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 3.2.16 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

<sup>35</sup>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  $\gamma = \gamma_0 \mathbb{1}_{N_S}$ .

matrices  $S_{50}$  instead of  $S_{25}$  **see if name is okay**. Looking at Figure [3.2.16](#) we see that points with  $\gamma_0 \gtrsim 0.7$  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  $\gamma_0, S_0, N_R$  and  $N_S$  such that we can work with a fairly high syntrophy, we still need to decide for which values of  $\alpha_0$  we compute  $\Delta_S^*(m, G, A)$ . For a fixed  $\gamma_0, S_0, G$  and  $A$ , we define the *critical dynamical syntrophy*  $\alpha_C^D(\gamma_0, S_0, G, A)$  as:

$$\alpha_C^D(\gamma_0, S_0, G, A) \equiv \max_{\alpha_0} \{ \alpha_0 : \mathcal{D}_{L,1}((\gamma_0, S_0, \alpha_0), G, A) = 1 \}. \quad (3.8)$$

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  $\Delta_S^*$  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  $\kappa_G$  and ecological overlap  $\eta_G$  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 **TO DO: change names here**:

- 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  $\alpha_C^D$  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  $\kappa_A$  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  $\Delta_S^*$  when there is no syntrophy at play. Figure 3.3.3 shows that structural stability confirms the trend hitherto observed: for a given ecological overlap  $\eta_G$ ,  $\Delta_S^*$  increases as the connectance  $\kappa_G$  increases and for a given  $\kappa_G$ ,  $\Delta_S^*$  decreases as  $\eta_G$  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<sup>36</sup>, 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  $\alpha_C^D$  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<sup>37</sup>. The only scenario where a clear effect is

<sup>36</sup>By “shuffling”, we mean that the  $l_\mu$  change but their average value does not. Indeed, recall that the  $\nu_\mu$  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.

<sup>37</sup>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.

observable is for  $\Delta_S^*$  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 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

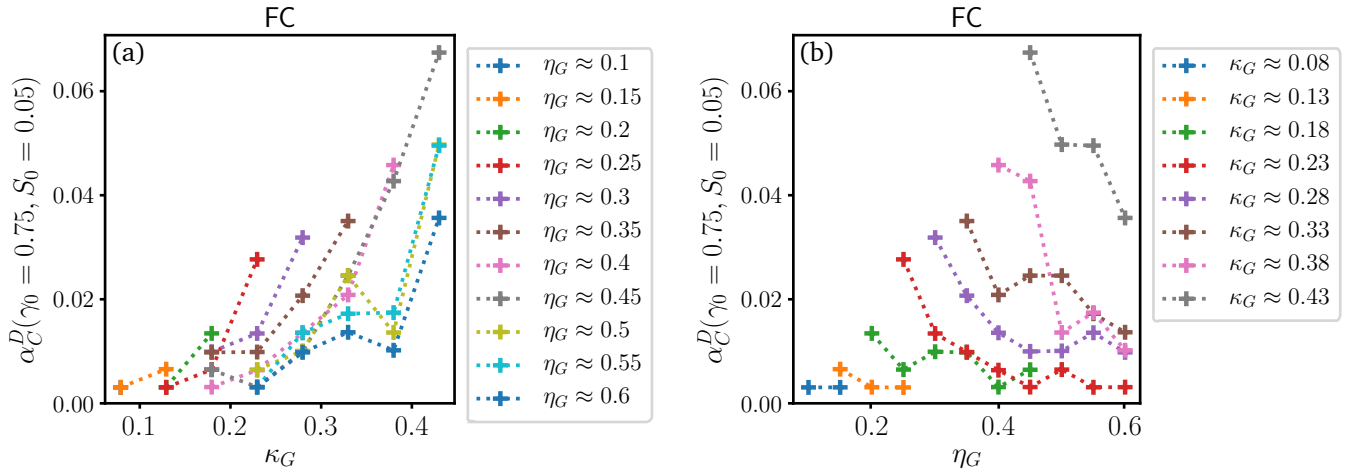


Figure 3.3.1: Critical dynamical syntrophy  $\alpha_C^D$  for all the consumption matrices  $G$  in the set  $S_{50}$  (fully connected syntrophy matrix) (a) as a function of the connectance of  $G$  for fixed ecological overlap and (b) as a function of the ecological overlap of  $G$  for fixed connectance. 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.

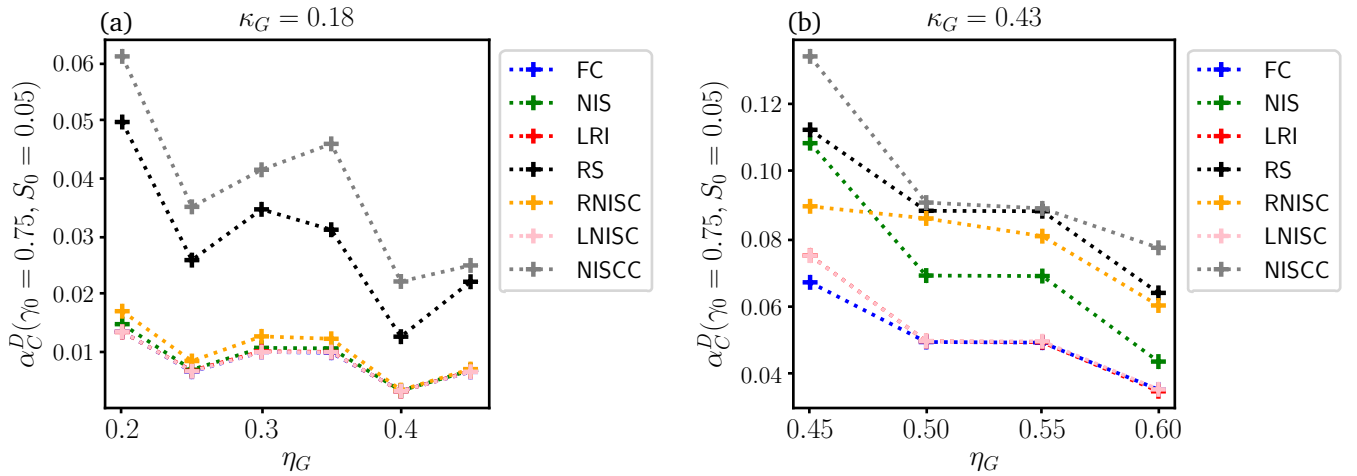


Figure 3.3.2: Critical dynamical syntrophy  $\alpha_C^D(\gamma_0 = 0.75, S_0 = 0.05, G, A)$  as a function of the ecological overlap of the consumption matrix  $G$  for fixed connectance (a)  $\kappa_G = 0.18$  and (b)  $\kappa_G = 0.43$ . The different lines symbolize the different structures of the 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??**

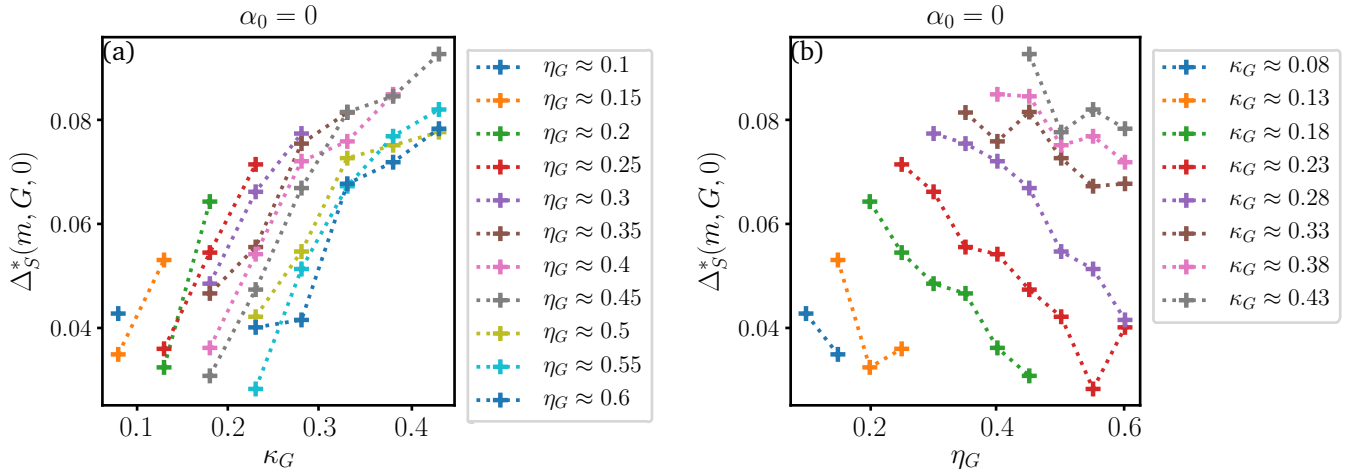


Figure 3.3.3: Critical structural perturbation without syntrophy  $\Delta_S^*(m, G, A = 0)$  (a) as a function of ecological overlap with fixed connectance and (b) as a function of connectance for a fixed ecological overlap. We look at matrices with  $N_R = 50$  and  $N_S = 25$  at one of the points in the metaparameters space that are the most dynamically stable for all the matrices (see Fig. 3.2.16), namely  $(\gamma_0, S_0) = (0.75, 0.05)$ . A clear trend may be observed, which is coherent with what was seen in Figure ???: 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.

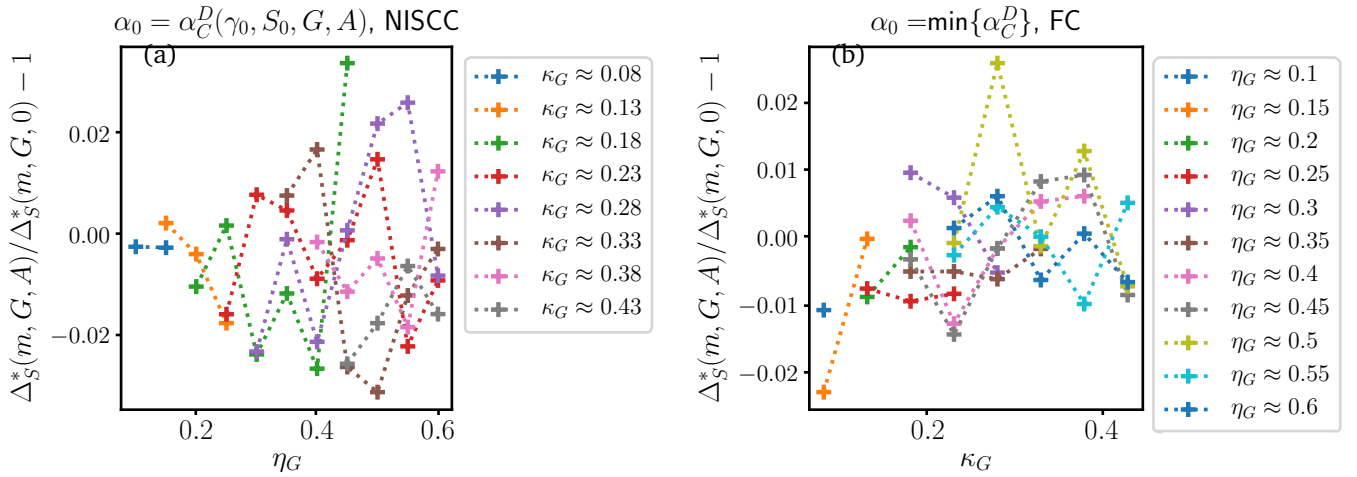


Figure 3.3.4: Typical deviation away from the "no syntrophy" scenario. For all structures of the syntrophy matrix considered, apart from NIS (see Fig. 3.3.5), adding a large syntrophy, be it the "common" or "individual" maximum, does not significantly increase the structural stability of the microbial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

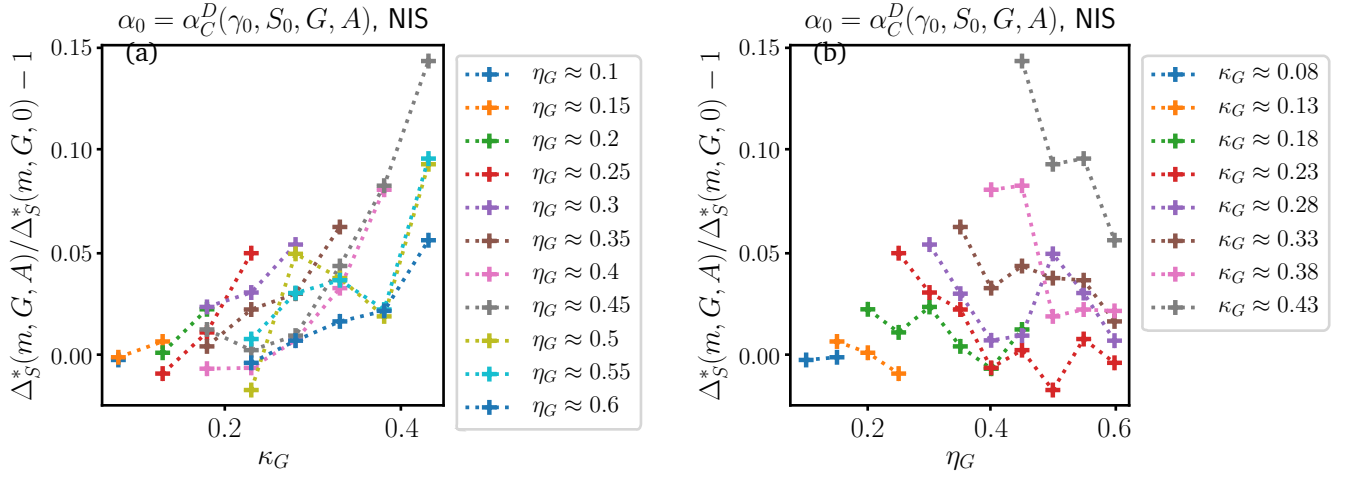


Figure 3.3.5: Deviation from the “no syntrophy” case for the syntrophy matrix  $A$  with a NIS structure.  $\alpha_0$  is taken at the critical dynamical syntrophy  $\alpha_C^D$  for each system. At the largest possible syntrophy,

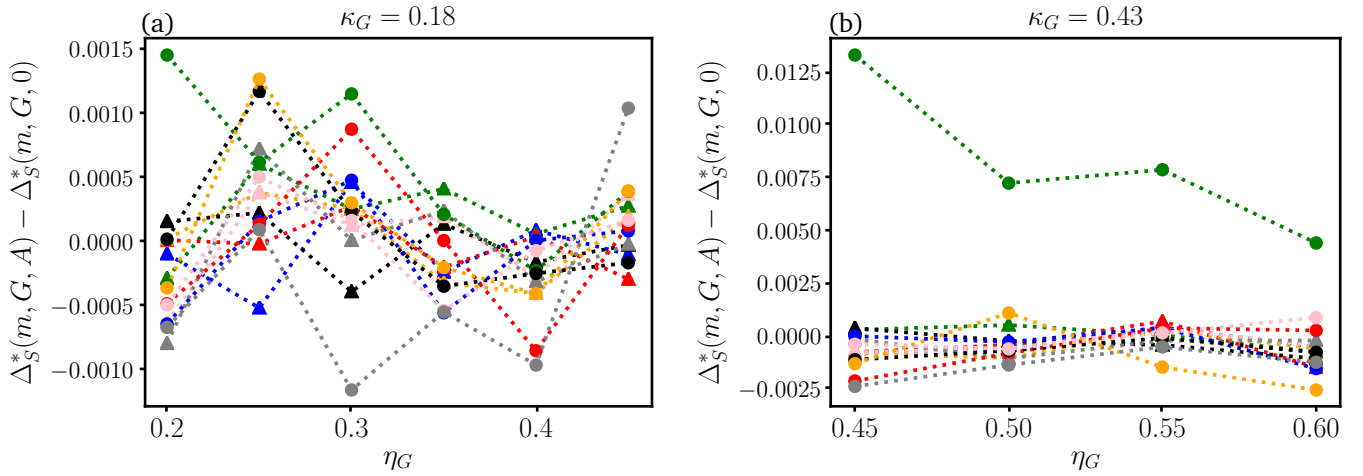


Figure 3.3.6: Difference between the critical structural perturbation with and without syntrophy for different scenarios as a function of the ecological overlap for a given connectance (a)  $\kappa_G = 0.18$  and (b)  $\kappa_G = 0.43$ . The different markers correspond to different  $\alpha_0$  taken : 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

The first significant result of this Thesis was the derivation of the equations that describe the metaparameters space domain in which feasibility is insured. This allowed us to accurately predict how syntrophy affects feasibility of microbial communities and told us the relevant zone of metaparameters (Methods [2.1.2](#)), which prevented us from dealing with unfeasible systems. We then discovered a special regime of parameters, named “strong LRI regime” (Methods [2.2.4](#)), which we proved is dynamically stable. This led to the development of the MCMC “LRI” algorithm explained in Methods [2.2.4](#) which goal is, for a given set of metaparameters and a consumption matrix, to generate the syntrophy matrix that brings microbial communities as close as possible to the strong LRI regime. Typically the syntrophy matrices obtained through that algorithmic procedure are very nested, such that the species which eat few resources end up releasing a lot of them and vice-versa. However, computations in Methods [3.2.3](#) showed that the goal behind the LRI algorithm is only partially achieved: microbial communities in the LRI syntrophic scenario can bear a larger syntrophic interaction while remaining dynamically stable – compared to the “null case” of a fully connected syntrophy matrix – but only if the average number of resources eaten by each consumer is low. Why the LRI scenario is inefficient at high consumption matrix connectance is as of now largely unknown. It could be due to feasibility considerations. It is indeed still unclear in which conditions the strong LRI regime is feasible, and it may happen that when the connectance of the consumption matrix is too large, such a regime simply cannot exist<sup>[38](#)</sup>. It could also come from the choice of the energy function used in the MCMC LRI algorithm. That energy was indeed obtained in a heuristic way that neglected some factors, especially the role of the critical radius and its dependency on the shape of the syntrophy matrix, and could be bettered in a future work.

We subsequently focused our attention on studying fairly large communities comprised of twenty-five microbial species and the same number of resources. As predicted analytically, it was confirmed numerically that imposing the fulfillment of simple physical conditions restricts quite much the range of possible metaparameters: not all configurations have an equal chance of existing. Indeed in the absence of syntrophy (Results [3.1.1](#)) we observe for all consumption networks a trade-off between average consumption rate and abundance of consumers in the sense that the overall proportion of resources consumed by the microbial species is bounded: there are no regimes where a lot of microbes eat very aggressively. If we want to increase the consumption rate beyond a certain threshold, the abundance of consumers has to be reduced, and the other way around. Adding a syntrophic mean-field interaction<sup>[39](#)</sup> breaks the symmetry of that trade-off: whatever the shape of the consumption network, as syntrophy increases only communities where lowly abundant microbial species aggressively consume their resources can exist **TO DO : check if there is any literature on this**. Moreover, communities where the average number of resources eaten by a consumer is large, under the caveat that there should be a small consumption overlap, are the ones whose existence is least impacted by syntrophy **TO DO: check if there is literature saying that LV systems are less feasible when nestedness is**

---

<sup>38</sup>More specifically, the critical radius  $R_C$  strongly depends on the shape of the consumption matrix and gets larger as its connectance increases. It is possible that beyond a certain connectance,  $R_C$  gets too large and the inequality of Theorem [1](#) cannot be fulfilled anymore.

<sup>39</sup>That is a fully connected  $A$  matrix (FC scenario).

**high, rephrase this better if possible.**

Whereas the feasibility region offered a fairly simple topology – the metaparameters space was made of two very clear regions of either complete feasibility or complete unfeasibility separated by a narrow area of mixed feasibility – the dynamical stability region presented a fractured landscape made of many almost dynamically stable points. Surprisingly enough, a similar trend as feasibility was observed: irrespective of the topology of the consumption matrix, microbial communities which have at equilibrium only a scarce abundance of consumers which eat their resources aggressively are the least impacted by syntrophy in terms of their dynamical stability. In general, as for the case of feasibility, microbial communities with highly connected consumption matrices with few syntrophic overlap are the ones which dynamical stability region is least impacted by the addition of a syntrophic interaction. However the same tendency being observed does not signify that all points of the metaparameters space which are feasible also are dynamically stable. On the contrary, for no consumption-syntrophy network did we find that feasibility implied dynamical stability for *all points* of the metaparameters space. A detailed study of exactly which sectors of the metaparameters space give rise to feasible communities that are always dynamically stable remains to be conducted. Among the points which we know are dynamically stable we were able to determine analytically (and that was confirmed by simulations) that the points at the upper border of the dynamical stability region are the most dynamically stable – in the sense that they take the least time to recover from perturbations of their abundances. Another trade-off worth mentioning is the one between size of dynamical stability region and “the speed of recovery” following perturbations. Indeed as syntrophy is increased, the size of the dynamically stable region shrinks<sup>40</sup> but the average largest real part of the spectrum of the jacobians observed within the dynamically stable region **TO DO : see exactly how that max is computed** increases. In short when the syntrophic interaction is raised, there are fewer dynamically stable points, but their stability also grows. This means we expect microbial communities where a large syntrophic interaction is observed to be fairly fine-tuned.

The study of structural stability, aka the impact of environmental perturbations, formed the final part of this Thesis. Because of reasons explained in the text, we decided to move to a system where the number of resources are doubled and the number of consumers is fixed. We chose a highly dynamically stable point of the metaparameters space and started by computing the critical dynamical syntrophy – the largest syntrophy such that we observe dynamically stable parameters with a probability of one – of each consumption-syntrophy network. In conformity with the trend hitherto observed, consumption matrices with a large connectance and a small ecological overlap can achieve the largest critical dynamical syntrophy. We then compared how microbial communities react to environmental perturbations in three different cases: as a null case, without syntrophy, then at their own critical dynamical syntrophy and as a “fairer” point of comparison, at the minimal critical dynamical syntrophy (*i.e.* the largest syntrophic interaction which can be sustained by all the matrices we considered). In all cases, we observed that, as for dynamical stability, the communities with a highly connected consumption matrix with little ecological overlap can recover from the largest environmental perturbations, which allows us to draw the

---

<sup>40</sup>Among others, because of the feasibility requirements.

conclusion that as regard to the shape of the consumption matrix dynamical and structural stability go hand in hand. Moreover, we were surprised to observe that – except for one specific scenario – syntrophy does not significantly impact structural stability of microbial communities, whether it was at the critical dynamical syntrophy of each network or at the minimal critical dynamical syntrophy. The only scenario where adding a syntrophic interaction undoubtedly increased structural stability was if the syntrophy was at its largest possible value (critical dynamical syntrophy) and the topology of the syntrophy matrix is such that consumers release every resource except the ones they do not eat. That result is very surprising to us because we observed that a random structure of the syntrophy matrix allowed a larger critical dynamical syntrophy, *i.e.* a larger syntrophic interaction without a loss of dynamical stability. That fact hints at a trade-off between dynamical and structural stability in terms of the shape of the syntrophy network that deserves to be explored in future studies **TO DO: to be sure of this should measure largest eigenvalue at that specific point for the different scenarios.**

So the shape of the syntrophy matrix has an effect on the structural stability of the system but it also affects feasibility and dynamical stability. While it surprisingly does not affect much the shape of the fully feasible region, it changes each consumption network's feasibility in a non trivial way. Overall we do not observe any significant increase of feasibility<sup>41</sup> for the NIS regime compared to the FC case, except for one particular lowly connected and lowly nested consumption matrix and highly connected highly nested consumption matrices. On the other hand, the LRI scenario greatly improves feasibility for lowly connected consumption networks but does not provide any significant change at high connectance. Finally, a random structure of the syntrophy matrix (RS regime) consistently improves feasibility, the lower the connectance of the consumption matrix, the larger the improvement. The same remarks hold for dynamical stability. The common ground between these cases is that the syntrophy matrix is lowly connected. Even though the reasons behind that behaviour remain unclear this hints that syntrophy matrices with a lower connectance enable a larger tolerance towards the addition of a syntrophic interaction.

All the previous conclusions are drawn for a system where the number of consumers is equal to the number of resources. There is however little reason that such systems should be prevalent in nature – and in fact they are not **TO DO: insert ref for this** – which begs the question of the consequences of changing *e.g.* the number of resources while keeping the number of consumers fixed. We roughly investigated what happens when the number resources goes from twenty-five to fifty. Interestingly enough, this reduces the common fully feasible region such that the largest syntrophy that *all* consumption-syntrophy can attain is lessened. However we also observed that doubling the number of resources does not – on average – significantly change the feasibility decay rate of each matrix. These contrasting conclusions, which may be due to a change of the profile of the feasibility function across the consumption rate-consumers abundance, need more investigation **TO DO: check maybe an error in the code.** As for the case of dynamical stability, interestingly enough adding more resources smoothes out the fully dynamically stable region, which eliminates most of the almost dynamically stable points and in turn allows for a common fully dynamically stable region at syntrophies larger than zero. Although the dynamical

<sup>41</sup>By “increase of feasibility” we mean “decrease of the feasibility decay rate”.

stability decay rate is on average left unchanged, there are hints that stability is increased (*i.e.* the largest real part of the spectrum is increased) when the number of resources is augmented, which is in accordance with the classical ecology literature **TO DO : cite here**. In any case, the effect of the number of resources of a microbial community remains an interesting open topic.

Although some progress on the subject of the impact of syntrophy on microbial communities was made with this Thesis, many aspects of the problem remain obscure. The first next step is of course to confront the results we obtained with experimental data **TO DO : put ref here?**. For instance, multiple considerations explained above lead us to expect that syntrophic microbial communities in Nature should have a low abundance of consumers at equilibrium (compared to the abundance of resources) which consume their resources at a high rate, with a very connected consumption network with little ecological overlap: how well is that prediction matched with experiments? Also if that is accurate, could microbial models explain the same results through for example an evolutionary model which would allow to bring a more biological hindsight? On the opposite, if our predictions are not matched with experiments, how can we change the model to make it more accurate?

Apart from confrontation with data, future theoretical studies still have significant work to do. For instance, a more exhaustive study of how structural stability changes as a function of the consumption rate and the abundance of consumers would surely provide interesting results: are there zones of the metaparameters space that are much sensible to perturbations of the consumers and the resources but not to environmental changes? Are there regimes that can sustain any perturbation at any intensity? However, in our opinion future works in the microbial ecology field should put their focus on finding the exact order parameter of the syntrophy matrix topology that drives the stability of this model. **TO DO : rephrase this: a bit rough as a final sentence**

## 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:

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

The equation we want to solve is:

$$\det(A_N) = 0. \quad (5.2)$$

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

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

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

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

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

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

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

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

With a Laplace expansion one gets:

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

This means:

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

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

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

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

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

Inserting this in Eq.(5.4) yields:

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

### 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 three terms that do not have the same behaviour: one term is linear in  $S_0$  with a negative coefficient, another is linear with a positive coefficient and another is quadratic (also with a positive coefficient). So we need to compute the minimum value the sum of these three terms is and take  $S_0$  as the minimum we found. The consumers equilibrium abundance  $S_0^*$  that yields the minimum value is implicitly given by the condition:

$$\frac{d}{dS_0} \left[ 4N_R\sigma_0\gamma_0S_0(\alpha_0 - \gamma_0R_0) + \frac{l_0^2}{R_0^2} + \frac{2N_S\alpha_0S_0l_0}{R_0^2} + \frac{N_S^2\alpha_0^2S_0^2}{R_0^2} \right]_{S_0=S_0^*} = 0. \quad (5.12)$$

The enthusiastic reader sees that this is equivalent to:

$$S_0^* = \frac{R_0^2}{N_S^2\alpha_0^2} \left( 2N_R\sigma_0\gamma_0(\gamma_0R_0 - \alpha_0) - \frac{N_S\alpha_0l_0}{R_0^2} \right). \quad (5.13)$$

One checks really easily that this point is indeed a minimum in  $S_0$ . So if  $S_0 > S_0^*$  it should be decreased, and otherwise increased. For  $\alpha_0 \rightarrow 0$ ,  $S_0^* \rightarrow \infty$ , and for  $\alpha_0 \rightarrow \infty$ ,  $S_0^* \rightarrow 0^-$ , which means we expect to find the most dynamically stable points at large  $S_0$  for low syntrophy and at low  $S_0$  for large syntrophy.

## 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  $R_\nu^*$  and  $S_i^*$  :

$$R_\nu^* = \mathcal{R} \ \forall \nu = 1, \dots, N_R \text{ and } S_i^* = \mathcal{S} \ \forall i = 1, \dots, N_S, \quad (5.14)$$

where  $\mathcal{R}$  and  $\mathcal{S}$  are random variables coming from a distribution of mean equal to the corresponding metaparameter and relative standard deviation<sup>42</sup>  $\epsilon$ . In our simulations, we chose uniform distributions :

$$\mathcal{R} \sim \text{Unif}(R_0, R_0\epsilon) \text{ and } \mathcal{S} \sim \text{Unif}(S_0, S_0\epsilon). \quad (5.15)$$

2. The efficiency matrix  $\sigma_{i\nu}$  is then drawn similarly, from a distribution with average  $\sigma_0$ . In order to simplify the problem<sup>43</sup>, we take a zero-variance like Butler and O'Dwyer in [20], i.e. all species consume resources at the same global efficiency :

$$\sigma_{i\nu} = \sigma_0. \quad (5.16)$$

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  $\gamma$ , 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  $\gamma_0$  and relative standard deviation  $\epsilon$ :

$$\gamma_{i\nu} = \text{Unif}(\gamma_0, \gamma_0\epsilon) G_{i\nu}. \quad (5.17)$$

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

$$l_\mu = \text{Unif}(l_0, l_0\epsilon) \ \forall \mu = 1, \dots, N_R. \quad (5.18)$$

5. The last free parameters are the non-zero elements of the syntrophy matrix  $\alpha_{\nu i}$ , since the  $d_i$  and  $l_\mu$  are determined through the equations of evolution at equilibrium. The adjacency matrix  $A$  of  $\alpha$  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  $\alpha$  from a uniform distribution of mean  $\alpha_0$  and relative standard deviation  $\epsilon$ :

$$\alpha_{\nu i} = \text{Unif}(\alpha_0, \alpha_0\epsilon) A_{\nu i}. \quad (5.19)$$

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  $m_\nu$ . All the parameters of the model are now fully determined.

<sup>42</sup>By relative standard deviation, we mean the standard deviation measured in units of the average value.

<sup>43</sup>Indeed, we observed that in general introducing a non uniform  $\sigma_0$  adds an additional not needed layer of complexity.

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):

$$\det \begin{pmatrix} -\Delta & \Gamma \\ B & 0 \end{pmatrix} = 0 \quad (5.20)$$

Using the fact that  $\Delta$  is invertible, we can make use of the equality<sup>44</sup>:

$$\det \begin{pmatrix} -\Delta & \Gamma \\ B & 0 \end{pmatrix} = \det(-\Delta) \det(B\Delta^{-1}\Gamma). \quad (5.21)$$

Eq. (5.20) then becomes:

$$\det(B\Delta^{-1}\Gamma) = 0 \quad (5.22)$$

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

$$\boxed{0 \in \sigma(J^*) \iff B\Delta^{-1}\Gamma \text{ is not full rank.}} \quad (5.23)$$

But when is  $B\Delta^{-1}\Gamma$  not full rank? Sylvester rank inequality [47] states that:

$$\text{rank}(B\Delta^{-1}\Gamma) \geq \text{rank}(B) + \text{rank}(\Delta^{-1}\Gamma) - N_R. \quad (5.24)$$

Similarly,

$$\text{rank}(\Delta^{-1}\Gamma) \geq \text{rank}(\Delta^{-1}) + \text{rank}(\Gamma) - N_R = \text{rank}(\Gamma), \quad (5.25)$$

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

$$\text{rank}(\Delta^{-1}\Gamma) \leq \min \{ \text{rank}(\Delta^{-1}), \text{rank}(\Gamma) \} \implies \text{rank}(\Delta^{-1}\Gamma) \leq \text{rank}(\Gamma). \quad (5.26)$$

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

$$\text{rank}(B\Delta^{-1}\Gamma) + N_R \geq \text{rank}(B) + \text{rank}(\Gamma). \quad (5.27)$$

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

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

<sup>44</sup>This uses a formula which is trivially analogous to one found in [31].



*Proof.* We assume that 0 is in the spectrum of  $J^*$  and prove it leads to a contradiction. Because  $0 \in \sigma(J^*)$ , Eq. (5.23) implies  $B\Delta^{-1}\Gamma$  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 :

$$\text{rank}(B\Delta^{-1}\Gamma) < \min(N_R, N_S) = N_R. \quad (5.28)$$

This can be used as an upper bound for Eq. (5.27) :

$$2N_R > \text{rank}(B) + \text{rank}(\Gamma). \quad (5.29)$$

However, we also know that B and  $\Gamma$  are full rank, *i.e.*

$$\text{rank}(B) = \text{rank}(\Gamma) = N_R. \quad (5.30)$$

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.  $\square$

#### 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

$$(\Gamma B)_{\mu\mu} < - \sum_{\nu \neq \mu} |(\Gamma B)_{\mu\nu}| \quad \forall \mu, \quad (5.31)$$

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

*Proof.* We assume

$$(\Gamma B)_{\mu\mu} < - \sum_{\nu \neq \mu} |(\Gamma B)_{\mu\nu}| \quad \forall \mu. \quad (5.32)$$

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

$$(\Gamma B)_{\mu\mu} - \lambda^2 < - \sum_{\nu \neq \mu} |(\Gamma B)_{\mu\nu}| \quad \forall \mu. \quad (5.33)$$

Dividing Eq. (5.33) by  $\Delta_\mu$ , we get:

$$\frac{1}{\Delta_\mu} \left[ \left( \sum_i \Gamma_{\mu i} B_{i\mu} \right) - \lambda^2 \right] < - \sum_{\nu \neq \mu} \left| \frac{\sum_i \Gamma_{\mu i} B_{i\nu}}{\Delta_\mu} \right| \quad \forall \mu. \quad (5.34)$$

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

$$S_{\mu\mu} + \sum_{\nu \neq \mu} |S_{\mu\nu}| < 0 \quad \forall \mu. \quad (5.35)$$

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.  $\square$

### 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

$$\frac{dR_\mu}{dt} \approx 0, \forall \mu. \quad (5.36)$$

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

$$R_\mu \approx \frac{l_\mu + \sum_j \alpha_{\mu j} S_j}{m_\mu + \sum_k \gamma_{k\mu} S_k}. \quad (5.37)$$

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

$$\frac{dS_i}{dt} = \left( \sum_\nu \left( \frac{\sigma_{i\nu} \gamma_{i\nu} l_\nu}{m_\nu + \sum_k \gamma_{k\nu} S_k} - \alpha_{\nu i} \right) - d_i + \sum_{\nu j} \frac{\sigma_{i\nu} \gamma_{i\nu} \alpha_{\nu j}}{m_\nu + \sum_k \gamma_{k\nu} S_k} S_j \right) S_i. \quad (5.38)$$

This can be rewritten in a more compact way:

$$\frac{dS_i}{dt} = p_i(S) S_i + \sum_j M_{ij}(S) S_i S_j \quad (5.39)$$

with

$$p_i(S) = - \left( d_i + \sum_\nu \alpha_{\nu i} \right) + \sum_\nu \frac{\sigma_{i\nu} \gamma_{i\nu} l_\nu}{m_\nu + \sum_k \gamma_{k\nu} S_k} \text{ and } M_{ij}(S) = \sum_\nu \frac{\sigma_{i\nu} \gamma_{i\nu} \alpha_{\nu j}}{m_\nu + \sum_k \gamma_{k\nu} S_k}. \quad (5.40)$$

If we assume the species  $S_k$  are not too far away from their equilibrium values<sup>45</sup> i.e.

$$S_k \approx S_k^* \forall k, \quad (5.41)$$

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

$$m_\nu + \sum_k \gamma_{k\nu} S_k \approx m_\nu + \sum_k \gamma_{k\nu} S_k^* = \frac{l_\nu + \sum_k \alpha_{\nu k} S_k^*}{R_\nu^*} \quad (5.42)$$

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

$$p_i(S) \approx p_i \equiv - \left( d_i + \sum_\nu \alpha_{\nu i} \right) + \sum_\nu \frac{\sigma_{i\nu} \gamma_{i\nu} l_\nu R_\nu^*}{l_\nu + \sum_k \alpha_{\nu k} S_k^*}, \quad (5.43)$$

and

$$M_{ij}(S) \approx M_{ij} \equiv \sum_\nu \frac{\sigma_{i\nu} \gamma_{i\nu} R_\nu^* \alpha_{\nu j}}{l_\nu + \sum_k \alpha_{\nu k} S_k^*}. \quad (5.44)$$

<sup>45</sup>Note that this is very rarely true, especially in the context of the study of structural stability, where entire species sometimes die out.

**Perturbation analysis**

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

$$S = S^* + \Delta S, \text{ with } \Delta S \ll 1. \quad (5.45)$$

Written this way, the effective equations of motion Eq. (5.39) are equivalent to:

$$\frac{d\Delta S_i}{dt} = p_i(S^* + \Delta S) (S_i^* + \Delta S_i) + \sum_j M_{ij}(S^* + \Delta S) (S_i^* + \Delta S_i) (S_j^* + \Delta S_j). \quad (5.46)$$

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

$$\frac{d\Delta S_i}{dt} = \tilde{p}_i \Delta S_i + \sum_j E_{ij} \Delta S_j + \mathcal{O}(\Delta S^2), \quad (5.47)$$

with

$$\tilde{p}_i \equiv p_i(S^*) + \sum_k M_{ik}(S^*) S_k^*, \quad (5.48)$$

and

$$E_{ij} \equiv \left( \frac{\partial p_i}{\partial S_j} \Big|_{S^*} + M_{ij}(S^*) + \sum_k \frac{\partial M_{ik}}{\partial S_j} \Big|_{S^*} S_k^* \right) S_i^*. \quad (5.49)$$

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

$$p_i(S^*) = - \left( d_i + \sum_\nu \alpha_{\nu i} \right) + \sum_\nu \frac{\sigma_{i\nu} \gamma_{i\nu} l_\nu}{m_\nu + \sum_k \gamma_{k\nu} S_k^*} \quad (5.50)$$

and

$$M_{ik}(S^*) = \sum_\nu \frac{\sigma_{i\nu} \gamma_{i\nu} \alpha_{\nu j}}{m_\nu + \sum_k \gamma_{k\nu} S_k^*}. \quad (5.51)$$

Hence, using Eq. (5.48):

$$\tilde{p}_i = - \left( d_i + \sum_\nu \alpha_{\nu i} \right) + \sum_\nu \frac{\sigma_{i\nu} \gamma_{i\nu}}{m_\nu + \sum_k \gamma_{k\nu} S_k^*} \left( l_\nu + \sum_j \alpha_{\nu j} S_j^* \right). \quad (5.52)$$

This can be simplified using Eq. (5.42) and Eq. (1.2b):

$$\tilde{p}_i = -d_i + \sum_\nu \sigma_{i\nu} \gamma_{i\nu} R_\nu^* = \sum_\nu \alpha_{\nu i}. \quad (5.53)$$

With a similar computation, one finds

$$E_{ij} = \sum_\nu \frac{\sigma_{i\nu} \gamma_{i\nu} S_i^*}{m_\nu + \sum_k \gamma_{k\nu} S_k^*} (\alpha_{\nu j} - \gamma_{j\nu} R_\nu^*). \quad (5.54)$$

Finally, Eq. (5.47) can be recast in

$$\frac{d\Delta S_i}{dt} = \sum_j (J_E)_{ij} \Delta S_j, \quad (5.55)$$

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

$$(J_E)_{ij} = \sum_{\nu} \left[ \frac{\sigma_{i\nu} \gamma_{i\nu} S_i^*}{m_{\nu} + \sum_k \gamma_{k\nu} S_k^*} (\alpha_{\nu j} - \gamma_{j\nu} R_{\nu}^*) + \alpha_{\nu i} \delta_{ij} \right]. \quad (5.56)$$

We see that without surprise we find again the  $B, \Gamma$  and  $\Delta$  matrices coming from the jacobian at equilibrium:

$$(J_E)_{ij} = \sum_{\nu} \left[ \frac{B_{i\nu} \Gamma_{\nu j}}{\Delta_{\nu}} + \alpha_{\nu i} \delta_{ij} \right] \quad (5.57)$$

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  $\alpha_0$ . 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:

$$\text{Vol} \left( \mathcal{F}_1^{G,A}(\alpha_0) \right) \equiv \frac{\iint_{(\gamma_0, S_0) \in [0,1]^2} d\gamma_0 dS_0 \mathcal{F}(m, G, A)}{\iint_{(\gamma_0, S_0) \in [0,1]^2} d\gamma_0 dS_0} = \iint_{(\gamma_0, S_0) \in [0,1]^2} d\gamma_0 dS_0 \mathcal{F}((\gamma_0, S_0, \alpha_0), G, A). \quad (5.58)$$

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

$$\text{Vol} \left( \mathcal{D}_{L,1}^{G,A}(\alpha_0) \right) \equiv \iint_{(\gamma_0, S_0) \in [0,1]^2} d\gamma_0 dS_0 \mathcal{D}_L((\gamma_0, S_0, \alpha_0), G, A). \quad (5.59)$$

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  $\mathcal{D}_L$  as a function of  $\alpha_0$** : 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  $\alpha_0$ :

$$\text{Vol}(\mathcal{F}_1^{S_M}) \equiv \iint_{(\gamma_0, S_0) \in [0,1]^2} d\gamma_0 dS_0 \min_{(G,A) \in S_M} \{\mathcal{F}((\gamma_0, S_0, \alpha_0), G, A)\}. \quad (5.60)$$

In a completely analogous manner, the *common (local) dynamical stability volume* of a matrix set  $S_M$  is given by:

$$\text{Vol}(\mathcal{D}_{L,1}^{S_M}) \equiv \iint_{(\gamma_0, S_0) \in [0,1]^2} d\gamma_0 dS_0 \min_{(G,A) \in S_M} \{\mathcal{D}_L((\gamma_0, S_0, \alpha_0), G, A)\}. \quad (5.61)$$

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  $\{(\gamma_0^i, S_0^i) \text{ with } i = 1, \dots, N_{\text{points}}\}$  which are then summed up with their according weights, e.g. for Equation (5.58):

$$\text{Vol}(\mathcal{F}_1^{G,A}(\alpha_0)) \approx \frac{\sum_{i=1}^{N_{\text{points}}} \mathcal{F}((\gamma_0^i, S_0^i), \alpha_0, G, A)}{N_{\text{points}}}. \quad (5.62)$$

### 5.2.7 Center of dynamical stability

The general idea behind the center of dynamical stability ( $\langle \gamma_0 \rangle_D(\alpha_0), \langle S_0 \rangle_D(\alpha_0)$ ) is to show that as syntrophy increases, only systems with a large  $\gamma_0$  and a small  $S_0$  remain dynamically stable (and hence feasible). For the case of feasibility, that was an easy task, one only needs to look at a heat map of  $\mathcal{F}_1^{S_{25}}(\alpha_0)$  (Fig 3.1.8) and see that only such zones remain fully feasible as syntrophy grows. However, as explained in the main text, that figure can not be replicated for the case of dynamical stability. Because of the almost fully dynamically stable points, the fully dynamically stable region  $\mathcal{D}_{L,1}^{G,A}(\alpha_0)$  of each  $(G, A) \in S_{25}$  is very fragmented such that the common fully dynamically stable region  $\mathcal{D}_{L,1}^{S_{25}}(\alpha_0)$  vanishes for  $\alpha_0 > 0$  (Fig 3.2.5). But even though  $\mathcal{D}_{L,1}^{S_{25}}(\alpha_0)$  quickly vanishes, it does not mean that there are no dynamically stable  $(\gamma_0, S_0)$ , on the opposite (Fig 3.2.6)! This is the reason we need another measure to show how syntrophy changes the location of the dynamically stable points.

Intuitively, we would like something that tells us basically where *on average* the most dynamically stable  $(\gamma_0, S_0)$  are. This is very close to the physical idea of the center of gravity, which tells you where most of the mass of an object lies. In consequence we define the *center of dynamical stability* ( $\langle \gamma_0 \rangle_D(\alpha_0), \langle S_0 \rangle_D(\alpha_0)$ ) with the same formulae, namely:

$$\langle \gamma_0 \rangle_D(\alpha_0) \equiv \frac{1}{N_0^{S_{25}}} \sum_{(G,A) \in S_{25}} \frac{\iint_{(\gamma_0, S_0) \in \mathcal{D}_{L,0}^{G,A}(\alpha_0)} \gamma_0 \mathcal{D}_L((\gamma_0, S_0, \alpha_0), G, A)}{\iint_{(\gamma_0, S_0) \in \mathcal{D}_{L,0}^{G,A}(\alpha_0)} \mathcal{D}_L((\gamma_0, S_0, \alpha_0), G, A)} \quad (5.63)$$

and

$$\langle S_0 \rangle_D(\alpha_0) \equiv \frac{1}{N_{\hat{0}}^{S_{25}}} \sum_{(G,A) \in S_{25}} \frac{\iint_{(\gamma_0, S_0) \in \mathcal{D}_{L,\hat{0}}^{G,A}(\alpha_0)} S_0 \mathcal{D}_L((\gamma_0, S_0, \alpha_0), G, A)}{\iint_{(\gamma_0, S_0) \in \mathcal{D}_{L,\hat{0}}^{G,A}(\alpha_0)} \mathcal{D}_L((\gamma_0, S_0, \alpha_0), G, A)} \quad (5.64)$$

where  $\mathcal{D}_{L,\hat{0}}^{G,A}(\alpha_0)$  is the set of points where the dynamical stability function has mass (*i.e.* does not vanish):

$$\mathcal{D}_{L,\hat{0}}^{G,A}(\alpha_0) = \{(\gamma_0, S_0) : \mathcal{D}_L((\gamma_0, S_0, \alpha_0), G, A) > 0\} \quad (5.65)$$

and  $N_{\hat{0}}^{S_{25}}$  simply denotes the number of consumption-syntrophy networks  $(G, A) \in S_{25}$ , which have non-zero points, *i.e.* for which  $\mathcal{D}_{L,\hat{0}}^{G,A}(\alpha_0)$  is not empty.

### 5.2.8 Probability of dynamical stability when feasibility is ensured

We here define the *probability of being dynamically stable when feasible*  $\text{Prob}(\mathcal{D}_L|\mathcal{F})(G, A, \alpha_0)$  for a network  $(G, A)$  and a syntrophy  $\alpha_0$ . First consider a single point  $(\gamma_0, S_0, \alpha_0)$ . We use conditional probability theory [48] to get the probability  $\text{Prob}(\mathcal{D}_L|\mathcal{F})$  that  $(\gamma_0, S_0, \alpha_0)$  is dynamically stable if it feasible:

$$\text{Prob}(\mathcal{D}_L|\mathcal{F})(G, A, \gamma_0, S_0, \alpha_0) \equiv \frac{\text{Prob}(\mathcal{D}_L \cap \mathcal{F})(G, A, \gamma_0, S_0, \alpha_0)}{\text{Prob}(\mathcal{F})(G, A, \gamma_0, S_0, \alpha_0)}, \quad (5.66)$$

where  $\text{Prob}(\mathcal{D}_L \cap \mathcal{F})$ , resp.  $\text{Prob}(\mathcal{F})$  are the probabilities that  $(\gamma_0, S_0, \alpha_0)$  are dynamically stable and feasible, resp. feasible. Since we required feasibility in the definition of dynamical stability, we have  $\text{Prob}(\mathcal{D}_L \cap \mathcal{F}) = \text{Prob}(\mathcal{D}_L)$ . Using Eqs. (2.5) and (2.37), one gets from Eq. (5.66):

$$\text{Prob}(\mathcal{D}_L|\mathcal{F})(G, A, \gamma_0, S_0, \alpha_0) = \frac{\text{Prob}(\mathcal{D}_L)(G, A, \gamma_0, S_0, \alpha_0)}{\text{Prob}(\mathcal{F})(G, A, \gamma_0, S_0, \alpha_0)} = \frac{\mathcal{D}_L((\gamma_0, S_0, \alpha_0), G, A)}{\mathcal{F}((\gamma_0, S_0, \alpha_0), G, A)}. \quad (5.67)$$

Finally the probability of being dynamically stable when feasible is taken as the average over the feasible points  $(\gamma_0, S_0)$  of the previous quantity:

$$\text{Prob}(\mathcal{D}_L|\mathcal{F})(G, A, \alpha_0) \equiv \frac{\iint_{(\gamma_0, S_0) \in \mathcal{F}_{\hat{0}}^{G,A}(\alpha_0)} d\gamma_0 dS_0 \frac{\mathcal{D}_L((\gamma_0, S_0, \alpha_0), G, A)}{\mathcal{F}((\gamma_0, S_0, \alpha_0), G, A)}}{\iint_{(\gamma_0, S_0) \in \mathcal{F}_{\hat{0}}^{G,A}(\alpha_0)} d\gamma_0 dS_0} \quad (5.68)$$

where  $\mathcal{F}_{\hat{0}}^{G,A}(\alpha_0)$  is simply the set of not unfeasible points:

$$\mathcal{F}_{\hat{0}}^{G,A}(\alpha_0) \equiv \{(\gamma_0, S_0) : \mathcal{F}((\gamma_0, S_0, \alpha_0), G, A) > 0\}. \quad (5.69)$$

Note that in practice we do not take into account points outside of the unit square  $[0, 1]^2$  for obvious numerical considerations.

### 5.2.9 Estimate the critical structural perturbation $\Delta_S^*$

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

$$P_E(\Delta_S^*(m, G, A), m, G, A) = 0.5. \quad (5.70)$$

The most straight forward way to find  $\Delta_S^*$  is to solve Eq. (5.70) 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  $\Delta_S^*$  lies. We wrote a solver for Eq. (5.70) 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”  $\Delta_H$  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  $N_{\text{points}}$  points  $\Delta_S$  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:

$$S_f(\Delta_S) \equiv \frac{1}{1 + e^{-C_1(\Delta_S - C_2)}}, \quad (5.71)$$

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. (5.71), 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  $\Delta_S^*$ .

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  $\Delta_S$ , is estimated numerically through the following procedure:

1. Create a parameters set  $\mathcal{A}(m, G, A)$ .
2. Structurally perturb it by an amount  $\Delta_S$ .
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  $N_{\text{sys}}$  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  $N_{\text{points}}$  reduces the error on  $\Delta_S^*$  faster than increasing  $N_{\text{sys}}$ .

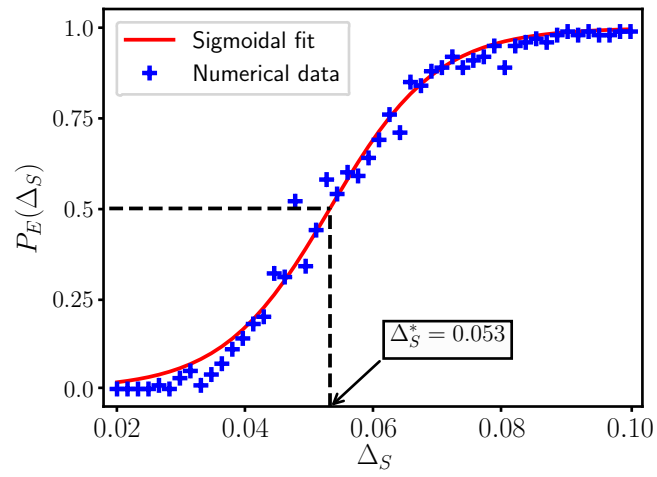


Figure 5.2.1: Typical probability of finding one extinction when structurally perturbing the system with a magnitude  $\Delta_S$ . The critical structural perturbation is easily estimated with a sigmoidal fit.



# Bibliography

- [1] Ashutosh Jogalekar. *Physicists in Biology; And Other Quirks of the Genomic Age*. en. Library Catalog: [blogs.scientificamerican.com](https://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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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. “Über 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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] Guy Bunin. “Ecological communities with Lotka-Volterra dynamics”. en. In: *Physical Review E* 95.4 (Apr. 2017), p. 042414. ISSN: 2470-0045, 2470-0053. DOI: [10.1103/PhysRevE.95.042414](https://doi.org/10.1103/PhysRevE.95.042414).
- [43] Y. Takeuchi. *Global dynamical properties of Lotka-Volterra systems*. en. Singapore ; River Edge, NJ: World Scientific, 1996. ISBN: 978-981-02-2471-4.
- [44] Peter Chesson. “MacArthur’s consumer-resource model”. en. In: *Theoretical Population Biology* 37.1 (Feb. 1990), pp. 26–38. ISSN: 00405809. DOI: [10.1016/0040-5809\(90\)90025-Q](https://doi.org/10.1016/0040-5809(90)90025-Q).
- [45] 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](https://doi.org/10.1103/PhysRevLett.118.048103).
- [46] 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](https://doi.org/10.1371/journal.pcbi.1006793).
- [47] 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](https://doi.org/10.1007/s00010-016-0412-4).
- [48] Andrey N. Kolmogorov. *Foundations of the Theory of Probability*. 1960.