Developing modified Lotka-Volterra models to simulate *in vitro* clonal dynamics

Author: Imanol Jurado Rodríguez*
Master en Física dels Sistemes Complexos i Biofísica.
Facultat de Física, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain.†

Advisor: Marta Ibañes Miguez
Department of Condensed Matter Physics and UBICS.
Facultat de Física, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain.

Advisor: Anna Alemany i Arias
Department of Anatomy and Embriology.
Leiden University Medical Centre Albinusdreef 2 2333 ZA Leiden, Netherlands.
(Dated: June 30, 2023)

Abstract: During the last decades, novel technological approaches have allowed uniquely labelling cells by integrating random barcodes in their DNAs. Such barcodes are permanent and are inherited by the cellular offspring. Thus, DNA sequencing permits their reading in order to identify those cells with common ancestors. This process, known as lineage tracing, enables the study of complex biological processes such as embryonic development, tissue homeostasis or even cancer metastasis with clonal, and even single cell, resolution. Here we aim to reveal the foundations behind the experimental results of clonal dynamics of colon cancer organoids through *in silico* simulations. We formulated three modified Lotka-Volterra models that allow us to investigate the role of clonal carrying capacity, proliferation rates and inter-clonal interaction network to achieve our purpose. The results show the vital role of partial interactions among clones and the importance of implementing nonequilibrium networks, *i.e.* architectures of interactions that vary in time. Furthermore, our results reveal a direct relationship between the harvesting time and the average number of surviving species at the end of the experiment, suggesting that external perturbations to the system can have big effects to clonal dynamics.

I. INTRODUCTION

The humankind has always had the eagerness to unveil the origin of our existence and to better comprehend the historical evolution that has driven the human race to present day. This is the reason why nowadays there are lots of researchers that still work on trying to reveal some of the inquiries about our progenitors. This ambition of knowing where we come from has also been transferred to a microscopic biological area. In fact, during the last decades there has been an increase of interest in enlightening the origin of all cells and their differentiation processes in our organisms. This procedure is named lineage tracing. It describes a set of methods that allows us to outline the fate of individual cells and their progeny [1]. Lineage tracing has been broadly applied to explore diverse complex biological processes including embryonic development, tissue homeostasis, stem cell functioning in regeneration and disease, and even cancer spreading [2–5].

A quarter of a century ago lineage tracing was performed following the cells by eye under the microscope during the embryonic development of *C. elegans* [6]. This animal has a deterministic development because all individuals of this species thrive practically in the same way. Moreover, the adult *C. elegans* is conformed by a relative small number of cells (~ 1000) and, since its body is transparent, it is possible to visualize them under the microscope. Thus, these factors simplify the study of their cellular division. Over time, more sophisticated technology allowed to mark cells with dyes or even with radioactive tracers to register cell positions, their divisions and their migratory tracks thanks to scanned laser light sheet fluorescence microscopy. These approaches have allowed to perform lineage tracing *in vivo* in zebrafish and mouse models [7]. Finally, during the last ten years, diverse tracing methods have been set up in which retroviral libraries are used to uniquely label single cells in a tissue with a permanent DNA barcode that is integrated in the cellular genome and inherited by the cellular offspring [8, 9]. By sequencing these barcodes, both within single cell and bulk resolution, lineage tracing can be performed.

These barcoding strategies allow to study clonal dynamics in a wide variety of systems both *in vivo* and *in vitro*. Here, we will focus on clonal dynamics in colon cancer organoids. Organoids are simplified 3D microscopic versions of an organ produced by cells extracted from a donor. Thus, organoids are capable to mimic the structure and to function similarly to the original organ [10]. The considerable advantage of organoids is that one can investigate in large quantities their performance under the effects of a disease and how they respond to possible treatments. Furthermore, they provide fast and robust

* imanol.juro99@gmail.com
† master.complex.biophys@ub.edu

outcomes because of their accurate representation of human tissue. Nowadays, there are organoid model systems for several organs, such as colon, lung, liver, brain, skin, among others [11–13].

In this project we intended to model *in silico* part of the experimental results obtained by Kester et al., 2022 [14]. The manuscript is organized as follows. In section II a description of the experimental set-up and the work by Kester and colleagues is summarized. Next, an overview of the Lotka-Volterra model and our proposed modified versions to model clonal dynamics in colon cancer are provided (section III). In section IV, linear stability analysis of our models is performed. In section V we investigate how inter-clonal interaction can be included in our models by means of an interaction matrix, and we investigate its effect in the average number of surviving clones. Next, we adapt our models to better simulate the experiments by using nonequilibrium networks to model organoid resampling (section VI). Finally, we discuss our results and propose future directions of research.

II. EXPERIMENTAL SET-UP AND PREVIOUS RESULTS

The cell lines used by Kester et al. were developed in Drost et al. 2015 [15], in which human intestinal stem cells from a healthy donor were manipulated genetically to incorporate mutations in some of their genes. Concretely, three cell lines were generated. In the first place, the APC gene, which acts as a tumor suppressor, was mutated to make it non-functional. The second cell line contains the APC mutation plus a mutation on P53, which is an essential gene to regulate cell proliferation. In the last cell line, another mutation was incorporated in KRAS, which controls the formation, maturation and destruction of cells. Drost and colleagues showed that each of the cell lines can generate colon organoids that recapitulate colon cancer phenotypes (such as appearence, proliferation rates and aneuploidy) at different stages of the disease, from the less lethal to the most aggressive one.

In Kester et al., the organoids were dissociated and, as a result, a dissolution of single stem cells was obtained (Fig.1). These single cells from each cell line were infected by a retrovirus which permitted to label each cell with a different barcode. The barcodes used have the following sequence of nucleotides: ATGCATGCATTTGTAAAACGACGGCCAGTNNNNNTNNNNNTNNNNNTNNNNNTCACACAGGAAACAGCTATGAGGCGCGCC, where A refers to adenine, T to thymine, C to cytosine, G to guanine, and N can be either A, T, C or G with equal probability. Since the barcodes have 20 N, that implies that theoretically there are 4^{20} of distinct barcodes and therefore we can uniquely label up to 4^{20} cells. Once all the cells were labeled, organoids were developed from every of them (Fig.1). From now on, since each organoid has its own barcode, we treat them as distinct species.

Arrived at this point, the system was let evolve periodically for one week during six months. At the end of every period, organoids were dissociated into single cells and a subset of them was used in order to quantify the proportions of every species by sequencing the barcodes of the cells. The rest of the cells were used to grow new organoids by depositing them again in a new Petri dish to repeat the process. Each repetion of this procedure is called passage and in every experiment a total of 20 passages were conducted. An example of the proportion of species recovered in every passage for one of the cell lines is displayed in Fig. 2, where we can observe the survival of only some of the original species. This outcome is counterintuitive because all families of organoids were initially supposed to be equal in the sense that there was no selective pressure, and their proliferative capacities were the same. However, Drost et al. observed that organoids accumulated aneuploidy (typical of cancer progression) over time, so that results may reflect this effect.

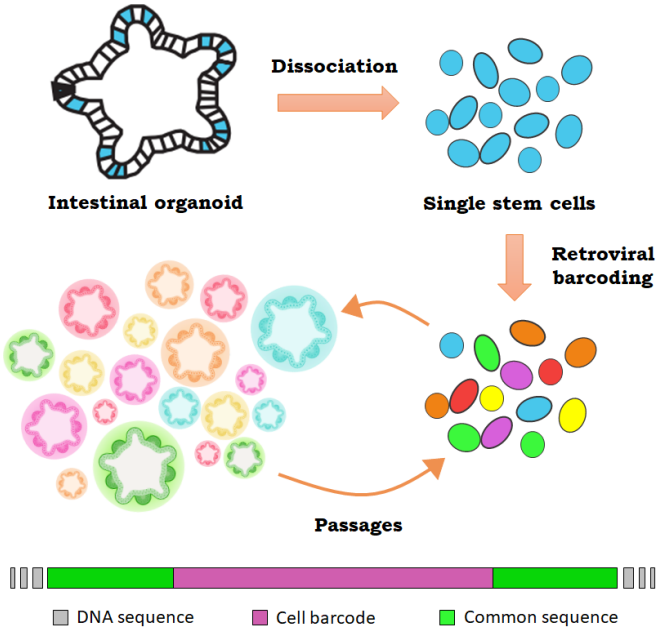

FIG. 1. *Cell barcoding and experimental procedure.* This illustration summarizes the experimental procedure conducted for the lineage tracing in the experiments done by Kester et al. [14], in which each colored cell/organoid indicates the presence of a different barcode. At the bottom a sketch of a cell barcode is shown. Common sequences are required for the detection of the cell barcode in the DNA, and the cell barcode is formed by 20 random nitrogenous bases.

Kester et al. computed a stochastic model without selective pressure in order to simulate *in silico* the behaviour observed in Fig. 2. In the simulations, identical species were let grow independently without considering any possible interaction among them. The effect of passaging was modeled as a Poisson sampling

acting as a bottleneck through which just a given sample of organoids was allowed to pass, representing the harvesting and relocation of a sample of organoids to another Petri dish. The results showed that a higher number of species should survive than the number observed experimentally. In addition, Kester et al. performed a second set of simulations in which initially each species had a different mutation and hence different replication rate. Here, they observed that the mutations with higher adaptability were the ones that survived. This means that before letting the system evolve, the surviving species could be predicted. However, this disagrees with what independent repetitions of the clonal dynamic experiments were showing.

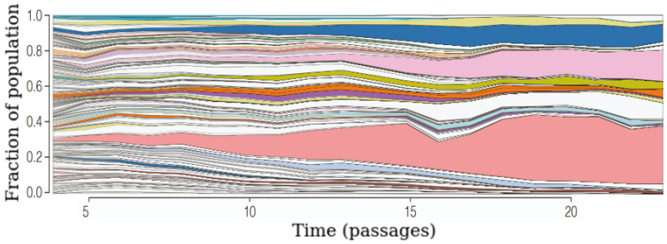

FIG. 2. *Experimental results of clonal dynamics.* This plot represents the *in vitro* clonal dynamics. The figure presents the relative frequency of observed viral lineage barcodes as a function of the number of passages. The image is courtesy of Kester et al., 2022 [14].

In front of these results, here we attempt to model *in silico* the clonal dynamics in order to unveil the main features behind the experimental phenomena observed. To do so, in this project we used a modified version of the very well-known Lotka-Volterra model [16, 17] to simulate the competitive dynamics of the populations of every species of organoids. Thus, one of the aims of this work is to demonstrate that an ecological model can also play an important role in the comprehension of the population dynamics of organoids and, eventually, organisms at microscopic level. We conclude that this model has projection because it could be applied to different biological processes such as cancer evolution, embryonic development or even cell organization in a tissue.

III. DEVELOPMENT OF MODIFIED LOTKA-VOLTERRA MODELS

Here we introduce the classical Lotka-Volterra model and our modified version of it. The following plots are outcomes of resolving the ordinary differential equations (ODEs) deterministically through the fourth-order Runge-Kutta (RK4) method [18]. One of the objectives of this research is to deal with multiple species, which implies the possible appearance of chaotic behaviour. Hence, high-order numeric methods are required in order to avoid generating artificial phenomenology. These methods have a considerable computational cost since they require more calculations and function evaluations. Nevertheless, a good balance is achieved with RK4 because accurate results are obtained without the necessity of a small step-size Δt. In fact, the time step chosen was of $\Delta t = 0.1$ in arbitrary units (a.u.) allowing fast computation without losing accuracy.

A. Logistic population growth model

Let us begin with a detailed description of the logistic population growth model for a single species. The first approach of intending to model population growth was conducted by Malthus in 1798 [19]. As can be seen in Eq.(1), he defined it as a birth-death process with constant rates:

$$\frac{dx}{dt} = bx - dx = (b-d)x = rx \Rightarrow x(t) = x_0 e^{rt} \qquad (1)$$

where x is the population size, $x_0 = x(t=0)$ is its initial value, b and d are positive constants representing the birth and death rates, and $r = b - d$ is the difference between both, which sets the net growth rate. The growth rate can be either positive or negative, so it gives rise to an exponential increase or decrease, respectively. It is important to emphasize that the population size x is the result of dividing the number of individuals by an arbitrary number, which could be a volume such that x could be interpreted as a density.

Although this approach may seem simple, there are regimes in the population growth of some species that follow this exponential behaviour. For instance, the growth of the humankind from 17th to 20th century was exponential, and it had projection to remain with the same rate for the 21st century [20]. However, human behaviour is complex and large repercussion phenomena can appear such as wars, diseases or lack of resources that can affect exponential growth. Evidently, these phenomena are not included in the Malthus model.

Thereafter, with the aim of correcting the unlimited exponential growth of the Malthus model, in 1845 the Belgium mathematician Verhlust proposed a self-limiting adjustment in order to model the overcrowding and limiting resources effects [21]. He suggested the following equation:

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{K}\right) \qquad (2)$$

being r the proliferation rate and K the so-called carrying capacity. Carrying capacity can be defined as the maximum population size of a biological species that can be sustained by a particular environment. It represents the limited amount of resources and the maximum stable population size that a species will tend to have. The term $-\frac{x^2}{K}$ models the interaction of the population with itself for the competition of resources.

As it is represented by the phase portrait in Fig. 3A, Eq.(2) has two fixed points that satisfy the stationary

condition $dx/dt = 0$, these are $x^* = 0$ and $x^* = K$. By means of a linear stability analysis we can demonstrate that $x^* = 0$ is an unstable fixed point, which means that any small deviation of the initial population size different from zero will grow. In fact, the dynamics will tend towards $x^* = K$; which is the stable fixed point. In case the initial population is higher than K, it will decrease until reaching the stable fixed point $x^* = K$ (Fig. 3B). Hence, the population size always approaches the carrying capacity value. Particularly, Eq.(2) can be solved analytically giving as a result [22]:

$$x(t) = \frac{K}{1 + Ce^{-rt}} \tag{3}$$

being $C = \frac{K - x_0}{x_0}$ and x_0 the initial population size.

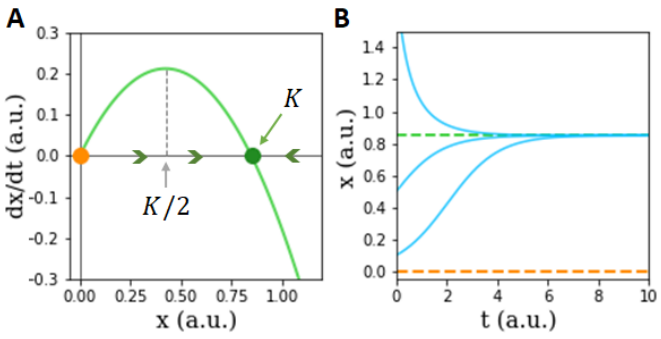

FIG. 3. *Logistic growth model.* **A**: Phase portrait of the logistic growth for $r = 1$ a.u. and a carrying capacity $K = 0.85$ a.u.. The two fixed points (orange for the unstable one and green for the stable one) are represented. The flow of the population size is displayed through the green arrows. **B**: Temporal evolution in arbitrary units of the population sizes $x(t)$ for different initial conditions. The dashed lines mark the fixed points $x^* = 0$ and $x^* = K$.

Further variations of the logistic model have been done with the intention of investigating certain biological behaviours observed in nature. To mention a few, the Allee effect to consider aggregation effects [23]. Ludwig proposed adding a nonlinear term representing the predation of budworms by birds to explain insect outbreak that caused the defoliation and death of a severe number of trees in the eastern North America forest [24]. Holling added a famous hyperbolic function, the Michalis-Menten equation with some adjustments, to model the harvesting and consumption of food [25]. In this project, we worked with the simplest situation which is the Verhlust logistic equation Eq.(2): this mathematical representation of the self-competition (also called self-interaction) for resources is the most similar to our experimental situation. Moreover, the value selected for the basal carrying capacity of all species was $K = 0.85$ a.u. based on Cho et al., 2022 [26], that describes the interaction of two types of cancer cells.

B. Lotka-Volterra equations for two interacting species

Here we introduce an extension of the Verhlust model by including the participation of a new species. This implies the addtion of an interaction term between the two species. This situation can be mathematically characterized through a well-known system of nonlinear differential equations called the Lotka-Volterra equations [16, 17]:

$$\begin{cases} \dfrac{dx}{dt} = r_x x \left(1 - \frac{x}{K_x} + \gamma_{xy} \frac{y}{K_x}\right) \\ \dfrac{dy}{dt} = r_y y \left(1 - \frac{y}{K_y} + \gamma_{yx} \frac{x}{K_y}\right) \end{cases} \tag{4}$$

being x and y the population size of each species, K_x and K_y their corresponding carrying capacities, respectively, and γ_{xy} and γ_{yx} are the parameters that define the type and the strength of the interactions between species. In fact, thanks to these interaction parameters, the Lotka-Volterra equations can be used in five different scenarios [27, 29]:

- Mutualism or symbiosis: both species take advantage of one another to grow themselves. It corresponds to the case in which both γ_{xy},$\gamma_{yx} > 0$. This type of interaction plays a crucial role in ecology by promoting and even maintaining the population of the species.
- Commensalism: one species benefits from the other without either harming or benefiting the latter. This situation is described by one positive interaction $\gamma_{xy} > 0$ and a null one $\gamma_{yx} = 0$.
- Predator-prey model: it is the most famous one. This situation is defined through antagonistic interactions, meaning that γ_{xy} and γ_{yx} are of opposite sign. This implies that one of the two species has the role of being food for the other one.
- Amensalism: this type of system reflects a biological interaction where one species, that remains unaffected, produces the destruction or inhibition of the other. It is mathematically expressed by one negative interaction $\gamma_{xy} < 0$ and a null one $\gamma_{yx} = 0$.
- Competitive model: in this situation both species compete for the same limited resources, thus one species inhibits the growth of the other one and viceversa. This system is described by both species having negative interactions γ_{xy},$\gamma_{yx} < 0$.

Once presented all the possible scenarios one can study, we can notice that the situation that concerns us is better described by the competitive model because we suppose that all species only compete for food. Furthermore, to simplify the study, the strength of the interactions was considered to be the same independently of the species,

$\gamma_{xy} = \gamma_{yx} = \gamma$. In addition, we also assumed that $r_x = r_y = r$ and $K_x = K_y = K$. In Fig. 4 we can observe different dynamics depending on whether the value of γ is higher, lower or equal than the self-interaction.

Firstly, looking at Fig. 4A, if the competition is weaker than the self-interaction, *i.e.* $-1 < \gamma < 0$, both species survive with the same population size, but with a smaller quantity than K.

Secondly, in Fig. 4B we observe the opposite situation. When the competition is stronger than the self-interaction, *i.e.* $\gamma < -1$, only one of the two species survives. The surviving species depends on the initial population sizes. This dichotomy is known as the *Principle of competitive exclusion*, which establishes that two species competing for the same limited resources cannot coexist.

Finally, Fig. 4C displays the situation in which the competition and the self-interaction are equal, *i.e.* $\gamma = -1$. In this case, we observe a line conformed of an infinite number of stable fixed points. Thus, extrapolating to multiple species, we expect a hypersurface of stable fixed points.

In the experimental set-up, we assumed that species interact among them in the same manner as with themselves (i.e. $\gamma = -1$). This enables to consider the case of all organoids (species) being equivalent and therefore we can set all the interactions to be the same.

C. Multispecies modified Lotka-Volterra models

Once expounded the foundations of the Lotka-Volterra equations, we can now infer the competitive model to multiple interacting species and formulate the following generalized equations:

$$\frac{dx_i(t)}{dt} = r_i x_i(t)\left(1 - \sum_{j=1}^{N} \frac{x_j(t)}{K_i}\right), \quad i = 1, ..., N \qquad (5)$$

where N is the total number of species, x_i is the population size of species i, r_i is its growth rate and K_i is its carrying capacity.

Returning back to the experimental data we aim to model, two of the main inquiries are: to what extent all species are actually equal? In case they are not, which factors should be considered to reproduce *in silico* the experimental results? With the aim of answering these questions, we must consider both biological and experimental conditions.

On the one hand, it is well-known that cancer cells proliferate without control and accumulate mutations giving them different fitness properties [28]. Thus, considering these species as cancer cells, a natural way to differentiate them is by means of modifying the growth rate r_i of each species. Then, we define an effective growth rate defined as $r_{\text{eff,i}} = r + \xi_i$, where r is a minimal basal growth rate shared between all the species, and ξ_i can take any

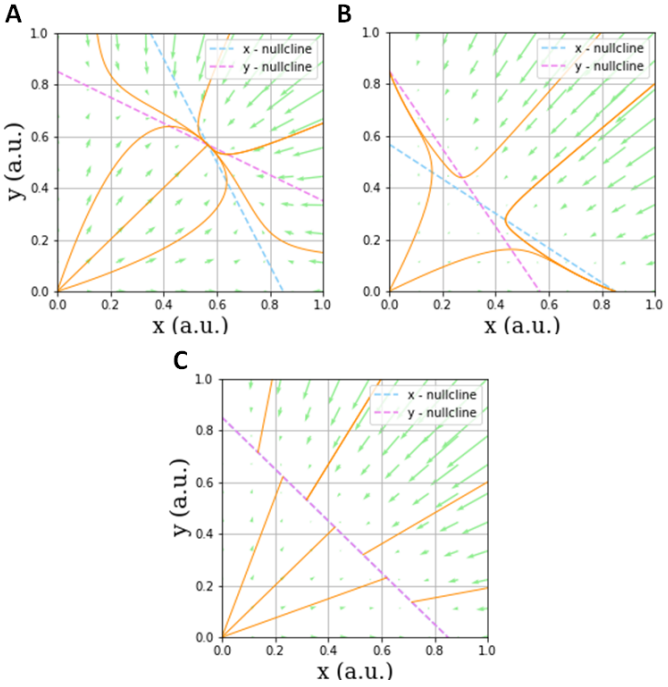

FIG. 4. *Phase portraits of the competitive model.* **A**: Phase portrait when the competition is weaker than the logistic self-interaction ($\gamma = -0.5$). Both species survive in the same proportion, but with a smaller size than K. **B**: Phase portrait when the competition is stronger than the logistic self-interaction ($\gamma = -1.5$). The *Principle of competitive exclusion* is observed. **C**: Phase portrait when both the competition and the logistic self-interaction are equal ($\gamma = -1$). There is a line of stable fixed points. In all plots $K = 0.85$ a.u. and $r = 1$ a.u.. The solid lines represent the trajectories of the dynamical system in the phase plane for different initial conditions. The dashed lines are the nullclines, meaning the solutions of $dx/dt = 0$ (blue) and $dy/dt = 0$ (violet). Finally, the green arrows are the vectorial representations of the trajectories.

positive value ($\xi_i \geq 0$). As a result, the modified version of the generalized Lotka-Volterra equations becomes:

$$\frac{dx_i(t)}{dt} = (r + \xi_i)x_i(t)\left(1 - \sum_{j=1}^{N} \frac{x_j(t)}{K}\right), \quad i = 1, ..., N \qquad (6)$$

where the carrying capacity is regarded the same for all species, meaning that all species compete for the same common resources. Therefore, we reduce the analysis to the essentials of having different effective growth rates $r_{\text{eff,i}}$.

In order to simplify the simulations, we divide the equations by r and re-scale the time through $\tau = rt$. Eq.(6) becomes then:

$$\frac{dx_i(\tau)}{d\tau} = (1 + \eta_i)x_i(\tau)\left(1 - \sum_{j=1}^{N} \frac{x_j(\tau)}{K}\right), \quad i = 1, ..., N \qquad (7)$$

being $\eta_i = \xi_i/r$.

On the other hand, another aspect that could be relevant in the dynamics is differentiating every species through the carrying capacity. Hence, we can deal with an effective carrying capacity $K_{\text{eff,i}} = K + \nu_i$, where K is the minimal basal carrying capacity due to the sources supplied in the experiment, and ν_i is also defined positive ($\nu_i \geq 0$). The term ν_i models the efficiency of every species to profit from the basal resources. In other words, if species i develops a certain feature that enhances its survival with respect the rest of species without affecting its proliferation rate, it will have a large ν_i as a result of this better adaptation. Then, we can express this new model as:

$$\frac{dx_i(\tau)}{d\tau} = x_i(\tau)\left(1 - \sum_{j=1}^{N} \frac{x_j(\tau)}{K+\nu_i}\right), \quad i = 1, ..., N \tag{8}$$

where the re-scaling of the time $\tau = rt$ has been maintained.

Finally, we also introduce a third model used as control in which all species are the same (*i.e.* $\xi_i = 0$ and $\nu_i = 0$). Then, the equations are written as:

$$\frac{dx_i(\tau)}{d\tau} = x_i(\tau)\left(1 - \sum_{j=1}^{N} \frac{x_j(\tau)}{K}\right), \quad i = 1, ..., N \tag{9}$$

To summarize, here we are going to compare the following three models:

- **Model A**, all species are exactly equal:

$$\frac{dx_i(\tau)}{d\tau} = x_i(\tau)\left(1 - \sum_{j=1}^{N} \frac{x_j(\tau)}{K}\right), \quad i = 1, ..., N \tag{10}$$

- **Model B**, every species has its own effective growth rate:

$$\frac{dx_i(\tau)}{d\tau} = (1+\eta_i)x_i(\tau)\left(1 - \sum_{j=1}^{N} \frac{x_j(\tau)}{K}\right), \quad i = 1, ..., N \tag{11}$$

- **Model C**, all species have different effective carrying capacity:

$$\frac{dx_i(\tau)}{d\tau} = x_i(\tau)\left(1 - \sum_{j=1}^{N} \frac{x_j(\tau)}{K+\nu_i}\right), \quad i = 1, ..., N \tag{12}$$

IV. LINEAR STABILITY ANALYSIS OF THE MODELS

With the purpose of understanding deeply the dynamics of the models, in this section we perform a linear stability analysis for each of them. In Fig. 5 we compare the analytical demonstrations with the results obtained through the simulations.

A. Model A: Equal species model

First, let us begin with the simplest situation where all species are equal (Eq.(10)). By solving $dx_i/d\tau = 0$, we find two set of possible solutions at the stationary situation:

$$\begin{cases} x_i^{st} = 0, & \forall i \\ \sum_i^N x_i^{st} = K \end{cases} \tag{13}$$

For the sake of simplicity, we are only going to study the homogeneous solutions, *i.e.* when all x_i^{st} are the same $\forall i$. This implies that either $x_i^{st} = 0$ or $x_i^{st} = K/N$ for $\forall i$. In order to know which of the solutions is stable, we apply a small perturbation to the steady state, $x_i = x_i^{st} + \delta x_i$ being δx_i small, and proceed with the stability analysis:

$$\frac{dx_i^{st}}{d\tau} + \frac{d\delta x_i}{d\tau} = (x_i^{st} + \delta x_i)\left(1 - \frac{Nx_i^{st}}{K} - \sum_j^N \frac{\delta x_j}{K}\right) =$$

$$= x_i^{st}\left(1 - \frac{Nx_i^{st}}{K}\right) - x_i^{st}\sum_j^N \frac{\delta x_j}{K} + \delta x_i\left(1 - \frac{Nx_i^{st}}{K}\right) \tag{14}$$

Higher-order terms are ignored so that we stay in the linear regime. Thereupon, by removing those terms that we already know are zero from the stationary solution of Eq.(10), Eq.(20) is reduced to:

$$\frac{d\delta x_i}{d\tau} = -x_i^{st}\sum_j^N \frac{\delta x_j}{K} + \delta x_i\left(1 - \frac{Nx_i^{st}}{K}\right) \tag{15}$$

Now, we can substitute both steady solutions and analyse the results.

For $x_i^{st} = 0\ \forall i$, Eq.(15) becomes:

$$\frac{d\delta x_i}{d\tau} = \delta x_i \tag{16}$$

Thus, a perturbation would grow exponentially fast, so the linear stability analysis indicates that $x_i^{st} = 0\ \forall i$ is an unstable solution.

For $x_i^{st} = \frac{K}{N}\ \forall i$, the expression obtained is:

$$\frac{d\delta x_i}{d\tau} = -\frac{1}{N}\sum_j^N \delta x_j \tag{17}$$

In order to achieve a better description of the stability of the fixed point, we must find the eigenvalues and eigenvectors that satisfy Eq.(17). Then, we rewrite Eq.(17) in its vectorial form:

$$\frac{1}{d\tau}\delta\overrightarrow{x} = A\delta\overrightarrow{x} \tag{18}$$

where $A = -\frac{1}{N}\overrightarrow{\overrightarrow{1}}$, being $\overrightarrow{\overrightarrow{1}}$ a matrix filled with ones.

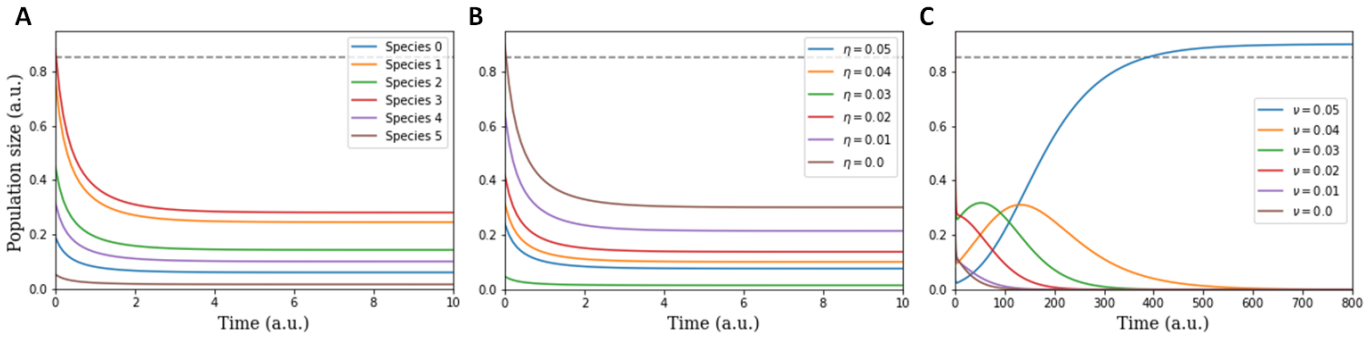

FIG. 5. *Dynamics of the modified competitive Lotka-Volterra models.* Here there are displayed the temporal evolutions (a.u.) of the population sizes (a.u.) of the three studied models for six species ($N = 6$): **A**: equal species model (Model A), **B**: effective growth rate model (Model B) and **C**: effective carrying capacity model (Model C). In every simulation, the interaction among six species has been computed. Indeed, their initial conditions are randomly selected through a uniform distribution U∈(0,1), but the steady solution would be the same if the initial conditions were equal. The dashed lines correspond to $K = 0.85$ a.u..

Now, the eigenvectors evolve over time as $\delta\overrightarrow{x}(\tau) = e^{\lambda\tau}\delta\overrightarrow{x}_s$, where $\delta\overrightarrow{x}_s$ is the spatial constant component. Then, we can seek the eigenvalues and eigenvectors:

$$\lambda\delta\overrightarrow{x}_s = A\delta\overrightarrow{x}_s \Rightarrow det(A - \lambda\mathbb{I}) = 0 \tag{19}$$

This shows that at least there is an eigenvalue $\lambda_0 = 0$, so there exists a N-dimensional hypersurface of solutions. And since the trace of A is negative, the solutions that fulfill $\sum_i^N x_i = K$ are classified as stable non-isolated fixed points [22]. As a matter of fact, this result was expected since in the 2D situation depicted in Fig. 4C we observed a line of stable solutions.

In Fig. 5A, we observe one of the infinite stable solutions this system can have for $N = 6$. In fact, in this figure the initial population sizes for each species have been selected randomly with a uniform probability distribution U∈(0,1). Hence, another initial condition would lead to another stable state.

B. Model B: Effective growth rate model

We proceed analogously with the linear stability analysis of the effective growth rate model. Therefore, we reach the same as in Eq.(15), but adding the $r_{\text{eff,i}}$ component:

$$\frac{d\delta x_i}{d\tau} = -r_{\text{eff,i}}x_i^{st}\sum_j^N\frac{\delta x_j}{K} + r_{\text{eff,i}}\delta x_i\left(1 - \frac{Nx_i^{st}}{K}\right) \tag{20}$$

Hence, if we consider again the homogeneous stationary solutions $x_i^{st} = 0$ and $x_i^{st} = \frac{K}{N}$ for $\forall i$, we find that the former is unstable. Let us now analyse the stability of the latter fixed point. As before, Eq.(20) becomes:

$$\frac{d\delta x_i}{d\tau} = -\frac{r_{\text{eff,i}}}{N}\sum_j^N\delta x_j, \quad \forall i \tag{21}$$

Thereby, we can rewrite the equations as in Eq.(18) where in this case:

$$A = \frac{1}{N}\begin{pmatrix} -r_{\text{eff,0}} & -r_{\text{eff,0}} & \dots & -r_{\text{eff,0}} \\ \vdots & \vdots & \ddots & \vdots \\ -r_{\text{eff,i}} & -r_{\text{eff,i}} & \dots & -r_{\text{eff,i}} \\ \vdots & \vdots & \ddots & \vdots \\ -r_{\text{eff,N}} & -r_{\text{eff,N}} & \dots & -r_{\text{eff,N}} \end{pmatrix} \tag{22}$$

It can be observed that every row can be obtained through a linear combination of the rest, this implies that $det(A) = 0$. Consequently, again at least a solution of $det(A - \lambda\mathbb{I}) = 0$ is $\lambda = 0$ and the trace is negative. Hence, for this model we have also found a N-dimensional hypersurface of stable non-isolated fixed points. We can argue that both the equal species model and the effective growth rate model behave practically in the same way. This is the reason why we observe similar quantitative dynamics comparing Fig. 5A and Fig. 5B.

C. Model C: Effective carrying capacity model

Here we compute the linear stability analysis for the effective carrying capacity model. First, by equaling to zero Eq.(12) we can infer the following possible stationary solutions:

$$\begin{cases} x_i^{st} = 0, & \forall i \\ \sum_j^N x_j^{st} = K_{\text{eff,i}} & \forall i \end{cases} \tag{23}$$

As done in the previous models, by means of substituting the stationary solutions we can discern their stability.

For $x_i^{st} = 0\ \forall i$, Eq.(25) is reduced to:

$$\frac{d\delta x_i}{d\tau} = \delta x_i \tag{24}$$

This solution is unstable because the perturbation would grow exponentially fast.

Now let us interpret the second constraint. It is indicating us that just one species survives with a population size $x_j = K_{\text{eff,i}}$, and the rest become extinct. This is fulfilled only when $j = i$. So, if we analyse another species, the only way to obtain the steady state $\frac{dx_m}{d\tau} = 0$ occurs when $x_m^{st} = 0$. Thence, we can deduce that there are N solutions: one for each species that satisfies $x_i^{st} = K_{\text{eff,i}}$ while $x_j^{st} = 0$ for $j \neq i$.

Thus, analogously to the previous analysis, we arrive at the following expression applying a small perturbation to Eq.(12):

$$\frac{d\delta x_i}{d\tau} = -x_i^{st} \sum_j^N \frac{\delta x_j}{K_{\text{eff,i}}} + \delta x_i \left(1 - \sum_j^N \frac{x_j^{st}}{K_{\text{eff,i}}}\right) \tag{25}$$

If we now look at the evolution of the perturbation of species m when $x_i^{st} = K_{\text{eff,i}}$ and $x_m^{st} = 0$ for $m \neq i$, Eq.(25) becomes:

$$\frac{d\delta x_m}{d\tau} = \left(1 - \frac{K_{\text{eff,i}}}{K_{\text{eff,m}}}\right) \delta x_m \tag{26}$$

Thus, the perturbation decays if $1 - \frac{K_{\text{eff,i}}}{K_{\text{eff,m}}} < 0$, so when $K_{\text{eff,i}} > K_{\text{eff,m}}$. Therefore, among the N solutions that satisfy $x_i^{st} = K_{\text{eff,i}}$, the only stable solution is the one in which only survives the species with largest $K_{\text{eff,i}}$.

These results are corroborated in Fig. 5C where we can see that only the species with the largest ν_i (meaning the largest $K_{\text{eff,i}}$) survives. Hence, we can conclude that for this model there is just one stable solution.

V. INTERACTION MATRIX

As we have observed in the previous sections, the interaction among species produces reciprocal effects in the dynamics of their populations. However, the experimental results present more heterogeneous dynamics and more variability in the final results, so a new inquiry emerges: where is the richness of the experimental dynamics originated?

Up to this point we have considered that all species interact with each other and with the resources in an identical manner. However, it is fundamental to describe the origin of the interactions. Despite this topic is unclear in the experiments, here we propose two types of scenarios:

- Fast resource diffusion: this interpretation lies on fast redistribution of the resources to the extent they are consumed. This implies that even if two species are localized far away from each other, the lack of supplies at a certain point will affect in the amount of food in another place and the competition will be global.

- Interaction by proximity: in this case we suppose that the resources are fixed in a lattice-like structure because we assume slow diffusion of food. Hence, only those species close to the same source compete among them.

These two hypothesis can fairly justify a global interaction among all species. Nevertheless, there could exist a more realistic framework in which we can better outline the experiments. The simplest and, at the same time, the most complex answer has to do with partial interactions among species. In other words, this means that every species just interacts directly with some other species. The reasoning behind this notion is related with the experimental set-up.

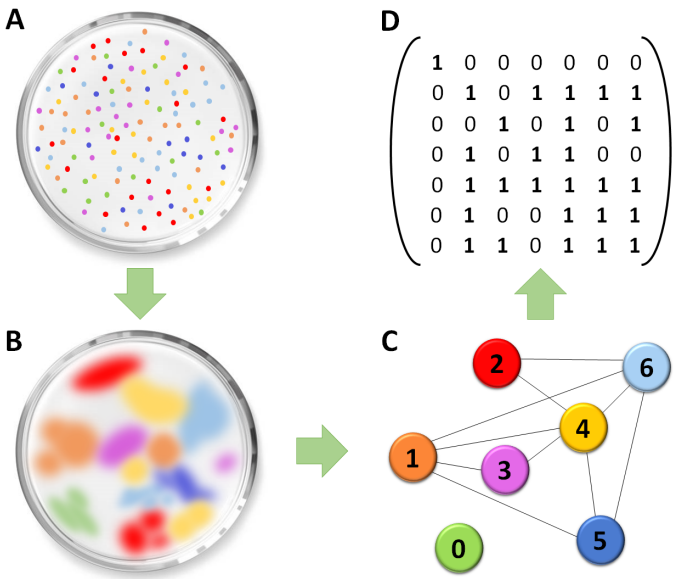

FIG. 6. *From the experimental set-up to the computational simulation.* This figure summarizes the interpretation given to the idea of partial interactions among seven species. **A**: Illustration of the initial moment in which labeled organoids are settled on the Petri dish (each dot represents an organoid). **B**: Sketch of the evolution of the system after a short time. The colored areas represent collections of organoids with the same label. Initial inhomogeneities become more notorious. **C**: Representation of the network of interactions extrapolated from **B**. **D**: Symmetric interaction matrix obtained from the sketched in **C**. This would be an example of a matrix that could be studied.

As described in section II, in the experiments, a collection of labeled organoids are deposited in a Petri dish with the proper medium for their development. This medium contains glucose, which is the food of the organoids. In order to decide which of the two hypothesis describes better the interaction among species, we calculated the time scale that glucose needs to diffuse along the culture. The diffusion coefficient of glucose in water at 24°C is of $D = 6.67 \cdot 10^{-10}\ m^2/s$ [30]. Thus, given the 5 cm of radius of a Petri dish, we can compute the diffusive time scale:

$$\tau_{glc} = \frac{L^2}{2D} \approx 3.1 \text{ weeks} \tag{27}$$

where $L = 5$ cm is the characteristic size of the Petri dish.

Since τ_{glc} is larger than the experimental harvesting period, we can consider that resources are placed in a lattice-like structure, so interaction by proximity hypothesis is plausible. Hence, we can induce a spatial description of the experiments and it is easier to implement the idea of partial interactions. To visualize this interpretation and reinforce the interaction by proximity hypothesis, Fig. 6A displays the supposed homogeneous initial distribution of organoids in the Petri dish. Nevertheless, there can exist particular inhomogeneities that become more notorious after letting the system evolve a certain time, as it is represented in Fig. 6B. In fact, the only way to understand this framework is by assuming interactions by proximity, where species that are in close proximity compete for the resources in that region.

The concept of partial interaction has been implemented by the so-called interaction matrix, namely γ_{ij}. The interaction matrix can be understood as a network of interactions as it is shown in Fig. 6C and Fig. 6D. Since the interactions are bidirectional due to the competitive model we have developed, this matrix is symmetric and is compounded by 1s and 0s that represent whether a species interacts or not with another one. Certainly, the underlying architecture of the interactions can be really complex and can form unique structures that, eventually, may generate peculiar emergent behaviours.

As a result of the implementation of this new concept, the original models (Eq.(10), Eq.(11), Eq.(12)) can be extended by adding the interaction matrix γ_{ij} as follows:

- Equal species model (Model A):

$$\frac{dx_i(\tau)}{d\tau} = x_i(\tau)\left(1 - \sum_{j=1}^{N}\gamma_{ij}\frac{x_j(\tau)}{K}\right), \quad i = 1, ..., N \tag{28}$$

- Effective growth rate model (Model B):

$$\frac{dx_i(\tau)}{d\tau} = (1+\eta_i)x_i(\tau)\left(1 - \sum_{j=1}^{N}\gamma_{ij}\frac{x_j(\tau)}{K}\right), \quad i = 1, ..., N \tag{29}$$

- Effective carrying capacity model (Model C):

$$\frac{dx_i(\tau)}{d\tau} = x_i(\tau)\left(1 - \sum_{j=1}^{N}\gamma_{ij}\frac{x_j(\tau)}{K+\nu_i}\right), \quad i = 1, ..., N \tag{30}$$

Here we attempt to observe patterns in the performance that species conduct for different shapes of the interaction matrix γ_{ij}. Thereby, to represent the experimental situation, we generated a collection of networks randomly connected using the Erdös-Renyi model [31]. One of the two ways to create this type of networks consists on considering a probability p to connect two nodes. Constructing the network in this way, if we increase p, we can generate the appearance of a giant component that percolates the system. In fact, the occurrence of this event undergoes a continuous phase transition when the probability of linking two nodes reaches a critical value $p = p_c$. Hence, once the network has been formed, we can define a probability of selecting randomly a node with k connections. This probability distribution is well defined through:

$$P(k) = \binom{N-1}{k} p_c^k (1-p_c)^{N-1-k} \tag{31}$$

where N is the number of nodes (species) in the network.

However, in this project we generated the random networks through a second method that also respects the probability distribution $P(k)$ and the properties of the Erdös-Renyi model. This method consists on fixing the total number of edges E we demand and, then, connecting pairs of nodes selected randomly until E edges are created.

With the new description of the models we cannot perform linear stability analysis, so from now on we will only work with simulations. In Fig. 7 we can observe the results obtained for the three models when $N = 6$ species interact among them with different connectivities. The connectivity is defined as the ratio between the number of edges E and the total possible edges the network can have, that is $N(N-1)$. The study of six species is just for practical reasons and to allow comprehensible figures for the reader because different number of species provides similar results. Additionally, whether the initial conditions were equal or not, the results obtained were the same. For this reason, we decided to compute all simulations with the same initial conditions ($x_i(0) = 1/N$ $\forall i$), so many more different networks could be studied. In total, each graphic contains $\sim$ 3720 simulations because a maximum of 300 different interaction matrices were computed for each connectivity. The arrival to the stationary state was determined by long enough simulations, that lasted up to 1000 time units.

Now, let us start by analysing the results depicted in Fig. 7A-C. The heatmaps present the probability of every species to survive as a function of the connectivity of the network. The top axis presents the average number of species that survive within every connectivity. Fig. 7A and Fig. 7B correspond to the equal species model and the effective growth rate model, respectively. We notice that both models behave similarly in relation with the surviving probability every species has: the more interactions there are, the less species survive; except for the situation in which all species interact. In this case all species survive as we proofed analytically. Furthermore, even when each species has a distinct effective growth rate, this does not produce any variation in the global behaviour. This result agrees with the analytical derivation done in the previous section in which we found that species with different growth rate behave analogously to species that are equal. Nonetheless, the results of the

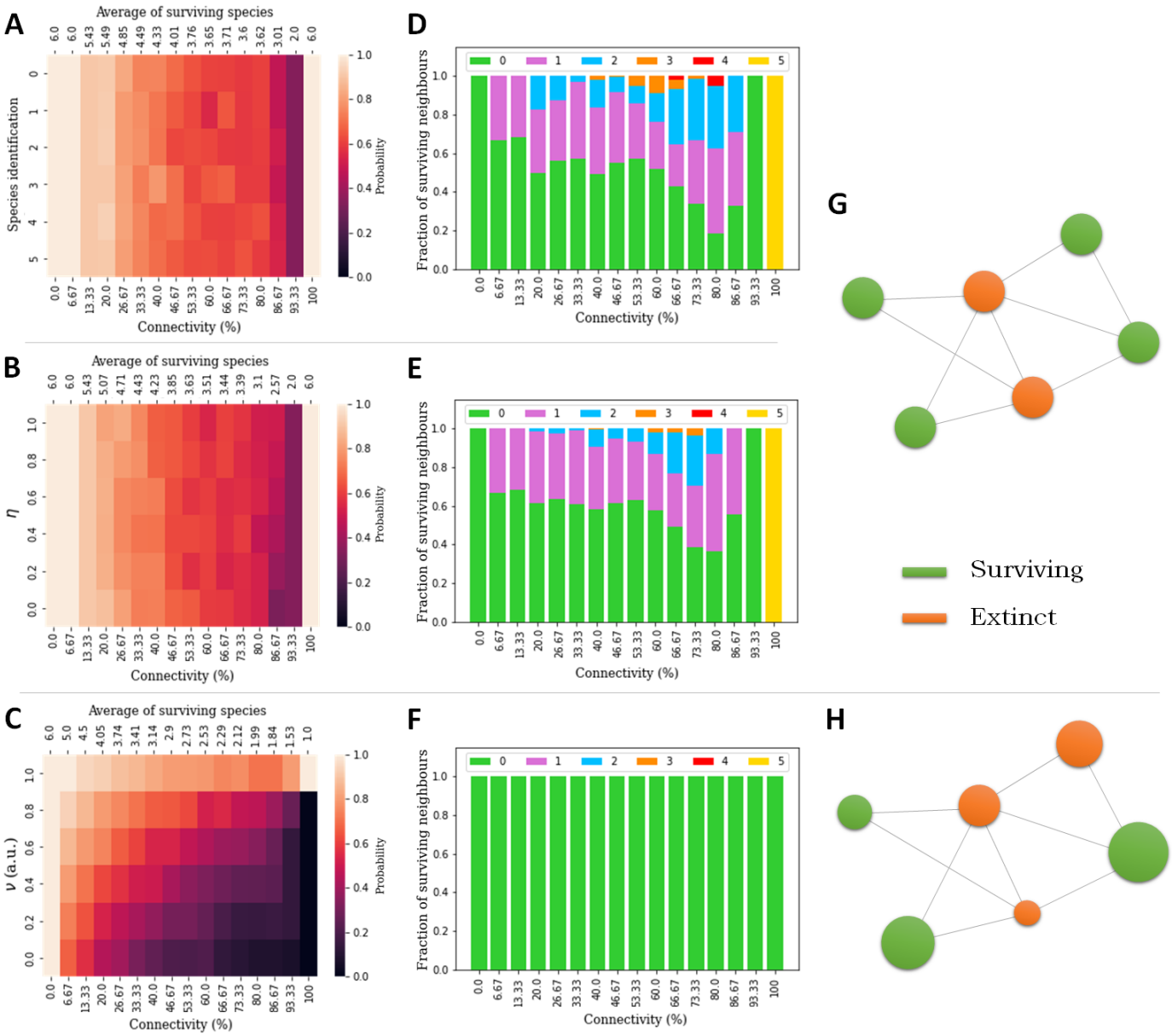

FIG. 7. *Surviving probability and topological pattern.* **A**, **B** and **C**: surviving probability as a function of the connectivity matrices for the equal species model (**A**), the effective growth rate model (**B**) and the effective carrying capacity model (**C**). The average number of surviving species for every connectivity is also shown. **D**, **E** and **F**: fraction of the number of surviving species as a function of the connectivity. The legend shows the number of surviving species that are first neighbours of a surviving species. **G** and **H**: illustrative examples of possible networks that have been studied, indicating which species survive (green dots) in the stationary state. In **H**, the size of the nodes is proportional to $K_{\text{eff,i}}$.

effective carrying capacity model in Fig. 7C show a different pattern. In addition to species being less likely to survive when the network is more connected, here we can also discern that the higher the effective carrying capacity compared with the other species, the more likely that species is to survive.

Arrived at this point, a question about the topological structure of the network arose: is there any topological pattern among the surviving species? Aiming to answer this question, we checked how many of the first nearest neighbours of a surviving species also survive. These outcomes are summarized in Fig. 7D-F. The barplots show the fraction of neighbours that survive around a surviving species as a function of the connectivity of the network. For instance, in Fig. 7G there are three surviving species with zero surviving neighbours, and two other surviving species that have one surviving neighbour. Thus, for this figure we would have $\frac{3}{5}$ with zero surviving neighbours, and $\frac{2}{5}$ with one surviving neighbour.

On the one hand, comparing the equal species model

with the effective growth rate model (Fig. 7D and Fig. 7E, respectively), we observe again similar results. The figures depict that there is a wide range of possible situations in which the first neighbours of a surviving species can also survive. An example of this case has been represented in Fig. 7G.

On the other hand, looking now at the results of the effective carrying capacity model in Fig. 7F, we can see that for any connectivity of the network every surviving species will be surrounded by species that have become extinct. This strong and powerful statement denotes a topological pattern that has been illustrated in Fig. 7H. Furthermore, this result holds for networks with more species (see Appendix C). Since topology and geometry are related, this topological pattern could allow us to extrapolate the position of each species in the Petri dish. This means that this study has also projection in the understanding of how cells organize spatially in a tissue.

VI. NONEQUILIBRIUM NETWORKS

In the experiments, once the populations of organoids have developed for one week, a sample from the culture is used to quantify the amount of cells of each species and deposited again in another Petri dish to continue its evolution. This process called resampling is done a total of twenty times. For this reason, we aimed to introduce the effect of resampling into the simulations.

In the previous section we have studied the dynamics of species that interact by proximity using a topological description given by networks. These networks have been generated through the Erdös-Renyi model and have been classified according to their connectivity. These type of networks are called equilibrium networks because the number of nodes is fixed. With the purpose of obtaining a more realistic description of the experiments, here we propose to incorporate nonequilibrium networks.

Nonequilibrium networks have the characteristic that the number of nodes changes in time. The vast majority of studies that use nonequilibrium networks consist on adding nodes [32]. Nonetheless, our situation is the opposite, nodes disappear when species become extinct. Hence, once a species vanishes, all those species that were also interacting with it now interact among them, and the system is let evolve again. This rearrange of the interactions is done once after every time period T. With this periodical reorganization of the interactions we intended to incorporate the experimental resampling. Additionally, the effective carrying capacity was multiplied by two in order to represent the relocation of the sample to a new Petri dish where there is twice space and supplies. Therefore, by means of doing this procedure recursively with an initial number of twenty species, we obtained dynamical evolutions like the ones portrayed in Fig. 8. Comparing both figures, we can notice that for the carrying capacity model (Fig. 8B) the dynamics drive towards the predominance of a certain species. However, for the equal species model (Fig. 8A), there is no such predominance. This last result is considerably similar compared with the real dynamics observed in the experiments (Fig. 2) in which all species were equal.

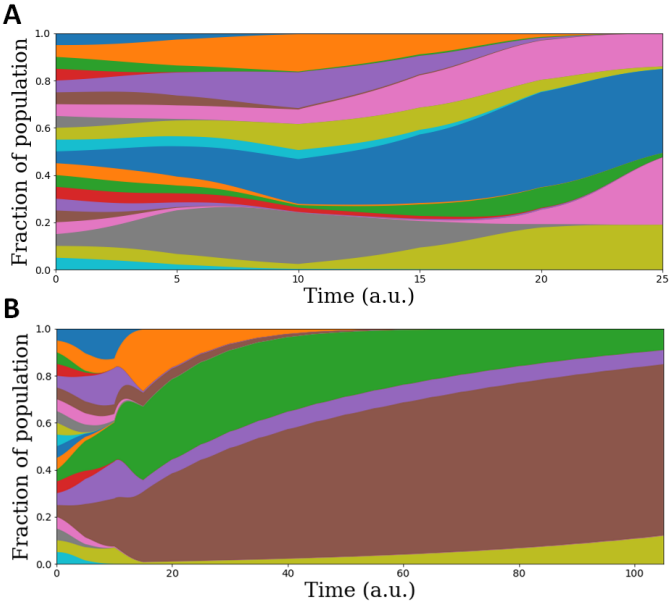

FIG. 8. *Result of simulating clonal dynamics using nonequilibrium networks.* The figures present the fraction of population of each species as a function of the time (a.u.). The plots depict the *in silico* clonal dynamics for the equal species model **A**, and for the carrying capacity model **B**. In both cases there is a total of 20 species that initially interact with a matrix of connectivity 41.58%. The harvesting time interval is of five time units.

Once we achieved representative results that resemble the experimental ones, we were interested in revealing some properties of the dynamics. Thence, we computed the probability that a species survives after the whole evolution when initially the network had a determined connectivity. We considered that a species became extinct if the fraction of its population dropped below 10^{-3} a.u.. Fig. 9A shows the survival probability for the equal species model. It is important to emphasize that the time interval T to reconnect the network was arbitrarily chosen to be of twenty time units and that, besides, a total of $\sim$28200 simulations for $N = 20$ species were performed.

As we would expect, in Fig. 9A we can see that at every initial connectivity the possibility of each species to survive is practically the same. Fig. 9B shows the average number of surviving species as a function of the initial connectivity of the interactions. For situations where the network is barely connected (from 0% to $\sim$17%), one can see a sharp decrease of the number of surviving species. Once this threshold is surmounted, the number of surviving species becomes quite independent of the initial connectivity, being on the range of five surviving species. This holds except for 100% of connectivity where all species survive as was shown analytically through the

linear stability analysis.

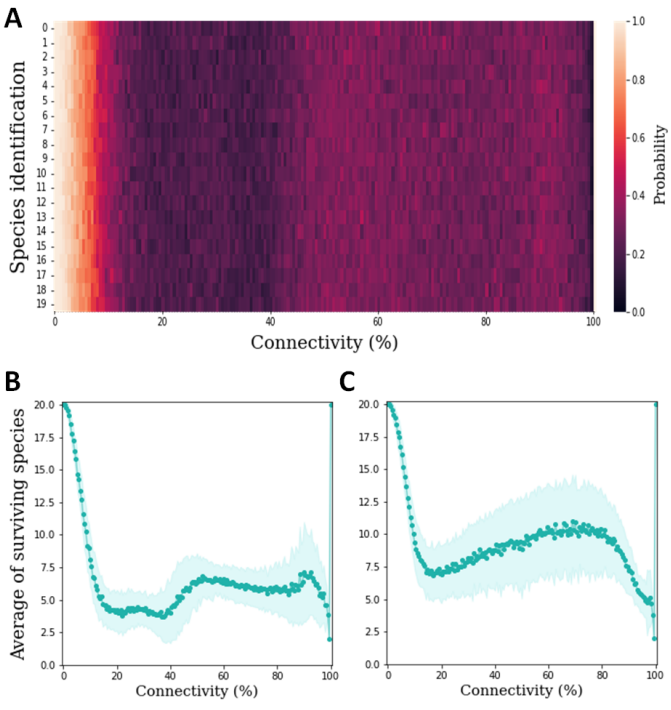

FIG. 9. *Analysis of survival for the equal species model.* **A**: Heatmap that displays the survival probability of each species as a function of the initial connectivity of the interaction matrix. The data was obtained for a time interval between gatherings of $T = 20$ time units. **B** and **C**: Average number of surviving species versus the initial connectivity when the time interval T is of 20 and 5 time units, respectively. The blue shades denote the standard deviations of the results.

Up to this point we have ignored the fact that we chose arbitrarily a time interval T for the reconnection of the interactions. Nonetheless, this gives rise to an important question: does the number of surviving species depend on the harvesting period of organoids? The origin of this question has to do with the fact that we might be interested in controlling the number of surviving species. In order to answer this question we repeated the same simulations but with a time interval T of five time units (Fig. 9C). Thus, we notice that when the ∼17% connectivity threshold is surpassed, the average number of surviving species rises up to ten. In other words, if we shorten the harvesting time interval, we are restraining the evolution of the system. This could explain why we observe in Fig. 2 that so many species survive. In fact, in the experiments the harvesting period was of one week, so we can extrapolate $T = 5$ time units from the simulations to $T = 7$ days from the experiments. Then, we can approximate the cell division rate as $r_{div} = \frac{r}{\ln 2}$, so:

$$r_{div} = \frac{1}{\ln 2} \cdot \frac{1 \text{ cell division}}{1 \text{ time unit}} \cdot \frac{5 \text{ time units}}{7 \text{ days}} \approx$$

$$\approx 1.03 \text{ cell divisions/day} \tag{32}$$

Thus, we can infer the time scale of cell division: $\frac{1}{r_{div}} \approx 0.97$ days, which matches the proliferating rate of human cells of ∼24h [33].

In the previous section we learnt that from the two models in which we differentiate every species either with the effective growth rate or with the effective carrying capacity, the latter is the one that provides patterns different from the reference model, *i.e.* the equal species one. Hence, since for the effective growth rate model we did not expect new results, we focused on repeating the same study we have discussed before, but for the effective carrying capacity model. We computed the simulations with a time interval of five time units to increase the average number of surviving species. However, with this model a new inquiry arises: has the relative mutations on environmental fitness an effect on the dynamics? This question can be rephrased as whether the difference between the effective carrying capacities of every species induces a distinct performance in the evolution. To try to answer this question we have to examine the results shown in Fig. 10.

Fig. 10A and Fig. 10B show the results when the difference between the effective carrying capacities of each species is large (*i.e.* the species with maximum $K_{\text{eff,i}}$ doubles the basal carrying capacity K). Looking at Fig. 10A, we can discern a regime in which the only surviving species are those with highest ν_i, *i.e.* those that are better adapted to the environmental conditions. Actually, at low initial connectivities one can perceive again the strong decrease of the average number of surviving species. Once the connectivity threshold of ∼17% is exceeded, the average number remains constant. This implies that the average number of surviving species is independent of the initial connectivity.

The results displayed in Fig. 10C and Fig. 10D correspond to small differences between the effective carrying capacities of each species. In fact, the highest ν_i is a fifth part of the basal carrying capacity K. By looking at Fig. 10C we note a substantial difference with the previous case (Fig. 10A). Now, there are more species that have a certain probability to survive, though the surviving probability of those species with largest ν_i is higher than the rest. Moreover, Fig. 10D shows that the average number of surviving species is higher compared to the case with larger differences among $K_{\text{eff,i}}$ (Fig. 10C). From this we can infer another relevant factor that determines the number of surviving species: the relative difference of the environmental fitness (*i.e.* the effective carrying capacity) among species.

VII. CONCLUSIONS

This project is a reflection of the process to develop a model to provide fundamental concepts to better understand clonal dynamics in colon cancer organoids. One of the aims of this project was to reproduce the *in vitro* experiments using the *in silico* simulations. The exper-

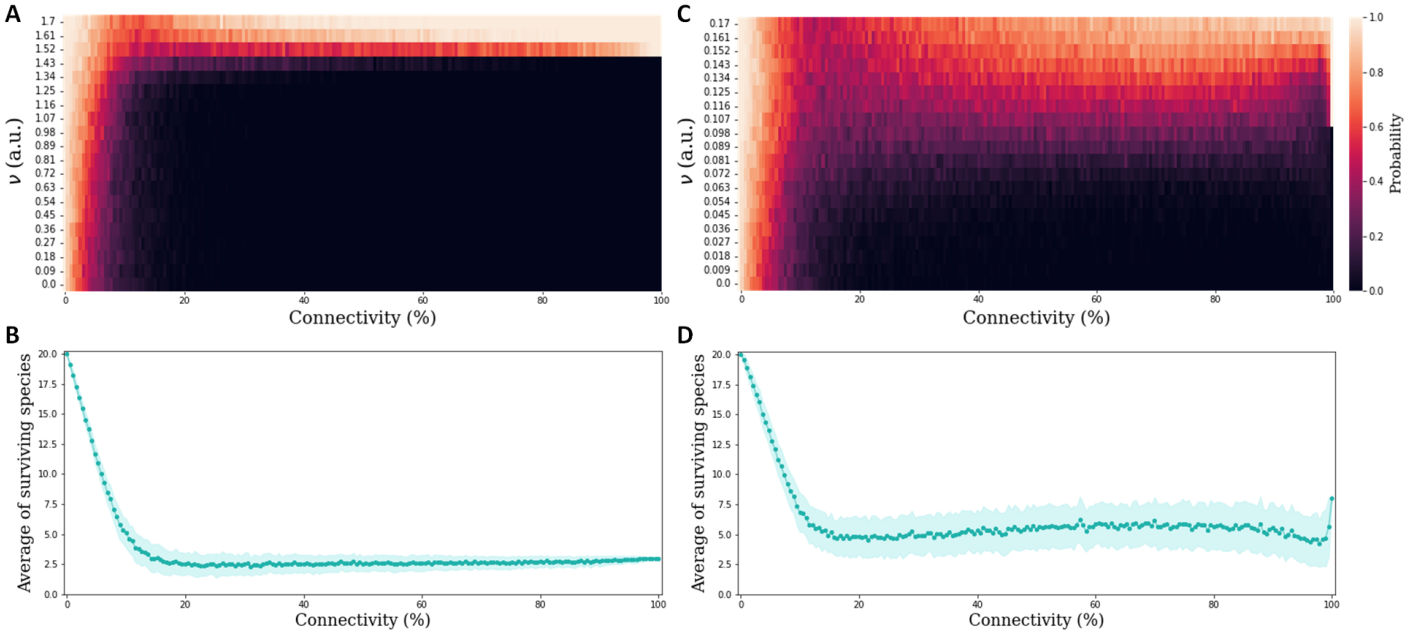

FIG. 10. *Analysing the survival of the effective carrying capacity model.* **A** and **C**: The heatmaps represent the surviving probability of every species differentiated by its own ν_i as a function of the initial connectivity of the interaction matrix. In the former the maximum is $\nu_i = 2K$, whereas in the latter the maximum is $\nu_i = K/5$. The data was obtained for a time interval between gatherings of 5 time units. **B** and **D**: These graphics present the average number of surviving species versus the connectivity of the initial interaction matrix for the two different cases mentioned before. The blue shades represent the standard deviations of the results

iments consisted on labeled organoids that were either equal or had different mutations and, in principle, all interacted with each other. Since the results exhibit the surviving of only certain species, we set out to investigate the vital factors behind this phenomenon. Comprehending the evolution and the characteristic properties of the model allow us to unveil hidden features or aspects that were ignored during the experiments. For these reasons simulations are so powerful; they help us identify the role of individual contributions to the phenomenon to desing better experimental plans.

Along this work we have used modified versions of the Lotka-Volterra model. This ecological model was originally created to study the population dynamics of interacting animals. However, here we wanted to investigate whether variations of this model could also serve for the microbiological world. This work shows the versatility of the model and that the results obtained could also be used to analyse the evolution of populations of cells. The decision behind the use of the Lotka-Volterra equations is that we were interested in simplifying as much as possible the model. Hence, we only needed the self-interaction given by the logistic growth and the competition among species. In this study we propose three models.

In the first situation, since all species were considered to be equal, both the self-interaction and the competition had the same value. Additionally, the growth rate and the carrying capacity of every species were also defined equal for all species. With these conditions we created the model that we named as the equal species model. This model present a N-dimensional hypersurface of possible stable solutions difficult to deal with.

Then, we were also interested in setting up a model in which every species was different. For this purpose, we built two models:

- The first model distinguishes every species through its proliferative capacity. This means that the multiplicative rate of every species is distinct. In fact, this could be a natural approximation to reflect cancer cells behaviour. However, this model and the equal species model provided similar outcomes, meaning that species that only differ in the proliferative capacity did not produce any substantial change in the performance. For this reason, we discarded this option and we worked with the equal species model instead.

- On the other hand, the other way to identify each species consisted on applying an effective carrying capacity for every species. This incorporates the concept of environmental fitness because it can be interpreted as the capacity of every species to favour its own development in comparison with the rest. This model generates different dynamics with respect to the equal species model. It displays a selective pattern in which just the species with the largest $K_{\text{eff,i}}$ survives.

Arrived at this point, we only had two scenarios: ei-

ther all species survive or just one survives. Hence, a crucial aspect that allowed us to obtain richer dynamics, a broader range of results and control the number of surviving species was the introduction of the interaction matrices. This important discovery revealed that there must be an spatial architecture of the interactions in order to achieve more realistic and similar performances to the experimental ones. In addition, we observed that the more interactions exist, the less species survive. For the effective carrying capacity model, those species with larger $K_{\text{eff,i}}$ are more likely to survive. This outcome is similar to the experiments in which all species had certain mutations. Indeed, those species with initial mutations that adapt better to the environment are the ones that survive. Hence, the model works properly to represent the experiments.

To comprehend deeper the structure formed in the topological networks, we analysed whether the first near neighbours of the surviving species became extinct or survived. For the effective carrying capacity model we noticed that every surviving species is isolated. This finding uncovers a relevant property in which two surviving species are those that do not interact directly.

Once this knowledge about our models was acquired, the following step consisted on adding more complexity to the dynamics that would allow us to get closer to the experimental results. To do so, we dealt with nonequilibrium networks that permitted us to integrate the loss of species in the interaction matrices. This step was decisive to accomplish our purpose of representing the clonal dynamics and accepting our models as good approaches. Moreover, these models enable us to observe similar features present in the experiments such as the extinction of the predominant population or even the almost extinction and the subsequent revival of a species. Certainly, the equal species model with the nonequilibrium interaction networks is an excellent approach to model the experiments. Additionally, this is supported when the time interval between harvestings is short, meaning that the dynamical evolution is constrained. Furthermore, in case every species has a subtle mutation, the effective carrying capacity model also works in the regime where the time between organoid harvests is short. With this model we have the advantage of knowing the characteristics of those species that have more probabilities to survive and the structural pattern they create. Moreover, we have observed that once the density of interactions surmounts the ∼17% threshold, the average number of surviving species becomes quite independent of the connectivity of the initial interaction matrix.

To summarize, from these models we have learnt that:

- The selective pressure arises from the competition and partial interactions among species.
- The environmental fitness represented by $K_{\text{eff,i}}$ allows to predict those species that have more probability to survive.
- In order to simulate more accurately the clonal dynamics, nonequilibrium networks are required.
- Reducing the period T between harvestings increases the average number of surviving species.

In the end, we have contemplated that these models work effectively to represent and better comprehend the foundations of clonal dynamics with the aim of performing lineage tracings. However, these models are so general that could also be applied to other medical fields like embryonic development, cancer metastasis or even to track the position of different cells in a tissue. The projection of the next study could be oriented to immunological inhibition of cancer cells in which some distinct species come from the immunologic system, whereas others species are cancer cells.

APPENDIX

A. Regular networks

During the process of understanding partial interactions in order to control the number of surviving species, we first dealt with a theoretical type of network that then would allow us to understand the patterns observed with random networks. This type of network is known as regular network. The reason behind is that we generated networks where all species had the same number of interactions. Let us compute the linear stability analysis for the equal species model for a regular network where all species interact only with another two. Thus, the equation we must to deal with is:

$$\frac{dx_i(\tau)}{d\tau} = x_i(\tau)\left(1 - \frac{x_{i-1} + x_i + x_{i+1}}{K}\right), \quad i = 1, ..., N \tag{33}$$

There are two types of stationary solutions:

$$\begin{cases} x_i^{st} = 0, \quad \forall i \\ x_{i-1}^{st} + x_i^{st} + x_{i+1}^{st} = K \quad \forall i \end{cases} \tag{34}$$

Again, to simplify the analysis we are only going to consider the homogeneous solutions meaning that $x_i^{st} = x_j^{st}\ \forall i \neq j$. Thus, the homogeneous solutions can be either $x_i^{st} = 0$ or $x_i^{st} = K/3\ \forall i$.

Now let us apply a small perturbation ignoring second order terms and those terms that are null:

$$\frac{d\delta x_i}{d\tau} = (x_i^{st} + d\delta x_i)\left(1 - \frac{3x_i^{st}}{K} - \frac{\delta x_{i-1} + \delta x_i + \delta x_{i+1}}{K}\right) =$$

$$= -x_i^{st}\frac{\delta x_{i-1} + \delta x_i + \delta x_{i+1}}{K} + \delta x_i\left(1 - \frac{3x_i^{st}}{K}\right) \tag{35}$$

Finally, we can substitute both homogeneous steady solutions and analyse the results:

- For $x_i^{st} = 0$:

$$\frac{d\delta x_i}{d\tau} = \delta x_i, \quad \forall i \tag{36}$$

Thus, a perturbation would grow exponentially fast, so the linear stability analysis indicates that $x_i^{st} = 0$ is the unstable solution.

- For $x_i^{st} = \frac{K}{3}$:

$$\frac{d\delta x_i}{d\tau} = -\frac{\delta x_{i-1} + \delta x_i + \delta x_{i+1}}{3}, \quad \forall i \tag{37}$$

In this case, as $\frac{d\delta x_i}{d\tau} < 0$, we can infer that a perturbation would decay exponentially fast. This means that $x_i^{st} = \frac{K}{3}$ is the stable solution.

Owing to the fact that the system presents translational invariance and all species are equal having the same number of connections, we propose the following eigenvector as a solution:

$$\delta x_j = \eta_s e^{i2\pi sj/N} e^{\lambda_s t}, \quad s = 1, ..., N \tag{38}$$

Thus, by substituting in Eq.(37) we obtain:

$$\lambda_s = -\frac{1}{3}\left(2\cos\frac{2\pi s}{N} + 1\right) \tag{39}$$

which shows that λ_s can be positive and therefore the uniform solution $x_i^{st} = K/3\ \forall i$ is unstable. We can observe that when $\frac{s}{N} = \frac{1}{2}$ corresponds to the maximum value of λ_s, which is the Fourier mode with fastest growth. This is the situation where every surviving species is surrounded by other that are extinct. Hence, this analysis predicts that surviving species will not be connected. However, this is for a specific network structure and connectivity. If the structure of the interactions is different there can appear other situations.

B. Network communities

Once discovered the worthy of using interaction networks, we studied the possibility of detecting the actual groups of survival and extinct species with network statistics. Concretely, we used the algorithm called Louvain method [34] to observe if the architecture of the network was enough to detect those communities that would survive. As a matter of fact, we included the whole equation into the interaction matrices as weights. Hence, we could detect the surviving species more efficiently without having to compute the whole dynamics. Nonetheless, we were restricted to analyse networks with few species, thus the statistics were confusing. Although for some networks the method detected the exact surviving group, in other cases it was wrong. However, it may work for larger networks, so it could be another branch of this study.

C. Fraction of surviving species $N = 20$

Fig. 11 shows the fraction of neighbours that survive around a surviving species as a function of the connectivity of the network for a total of $N = 20$ species. Although the number of species is increased, the same pattern is maintained.

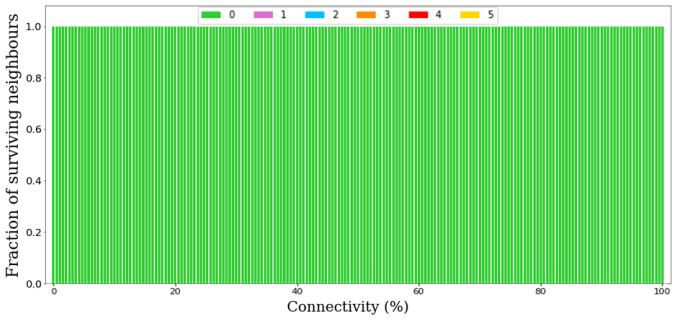

FIG. 11. *Fraction of surviving neighbours for* $N = 20$. The bar chart displays the fraction of the number of surviving species as a function of the connectivity for $N = 20$.

ACKNOWLEDGMENTS

I wish to show my appreciation to Dr. Marta Ibañes and Dr. Anna Alemany for their enthusiasm in guiding me throughout the project and for providing me some of their priceless knowledge. I would also like to thank my family and friends for all the support and advice.

[1] S. Van Horn and S. A. Morris, "Next-Generation Lineage Tracing and Fate Mapping to Interrogate Development", *Developmental Cell*, vol. 56, no. 1, pp. 7-21, 2021.

[2] K. Guan, J. Rohwedel, and A. M. Wobus, "Embryonic stem cell differentiation models: cardiogenesis, myogenesis, neurogenesis, epithelial and vascular smooth muscle cell differentiation in vitro", *Cytotechnology*, vol. 30, no. 1-3, pp. 211–226., 1999.

[3] Y. Zhang, F. Zeng, X. Han, J. Weng and Y. Gao, "Lineage tracing: technology tool for exploring the development, regeneration, and disease of the digestive system", *Stem. Cell. Res. Ther.*, vol. 11, no. 438, (2020).

[4] P. Rompolas, R. M. Kailin, K. Kawaguchi, P Sangbum, D. Gonzalez, S. Brown, J. Boucher, A. M. Klein and V. Greco, "Spatiotemporal coordination of stem cell commitment during epidermal homeostasis", *Science*, vol. 352, no. 6292, pp. 1471-1474, 2016.

[5] J. J. Quinn, M. G. Jones, R. A. Okimoto, S. Nanjo, M. M. Chan, N. Yosef, T. G. Bivona and J. S. Weissman,"Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts", *Science*, vol. 371, 2021.

[6] J. E. Sulston, E. Schierenberg, J. G. White and J. N. Thomson, "The embryonic cell lineage of the nematode Caenorhabditis elegans", *Dev. Biol.*, vol. 100, no. 1, pp. 64–119, 1983.

[7] P. J. Keller, A. D. Schmidt, J. Wittbrodt and E. H. Stelzer, "Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy", *Science*, vol. 322, no. 5904, pp. 1065-1069, 2008.

[8] S. F. Levy, J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher and G. Sherlock, "Quantitative evolutionary dynamics using high-resolution lineage tracking", *Nature*, vol. 519, no. 7542, pp. 181-186, 2015.

[9] M. G. Jones, D. Yang, and J. S. Weissman, "New Tools for Lineage Tracing in Cancer In Vivo", *AnnuRev-CancerBio*, vol. 7, pp. 111-129, 2023.

[10] Z. Zhao, X. Chen, A. M. Dowbaj et al., "Organoids", *Nat. Rev. Methods Primers*, vol. 2, no. 94, 2022.

[11] S. E. Blutt and M. K. Estes "Organoid models for infectious disease", *Annu. Rev. Med.*, vol. 73, pp. 167-182, 2022.

[12] J. Kim, B. K. Koo and J. A. Knoblich, "Human organoids: model systems for human biology and medicine", *Nat. Rev. Mol. Cell Biol.*, vol. 21, no. 10, pp. 571-584, 2020.

[13] N. S. Corsini and J. A. Knoblich, "Human organoids: New strategies and methods for analyzing human development and disease", *Cell*, vol. 185, no. 15, pp. 2756-2769, 2022.

[14] L. Kester, B. Barbanson, A. Lyubimova, L. Chen, V. van der Schrier, A. Alemany, D. Mooijman, J. Peterson-Maduro, J. Drost, J. Ridder and A. van Oudenaarden, "Integration of multiple lineage measurements from the same cell reconstructs parallel tumor evolution", *Cell Genomics*, vol. 2, no. 2, 2022.

[15] J. Drost, R.H. van Jaarsveld, B. Ponsioen, C. Zimberlin, R. van Boxtel, A. Buijs, N. Sachs, R.M. Overmeer, G.J. Offerhaus, H. Begthel, J. Korving, M. van de Wetering, G. Schwank et al., "Sequential cancer mutations in cultured human intestinal stem cells", *Nature*, vol. 521, no. 7550, pp. 43-47, 2015.

[16] A. J. Lotka, "Elements of physical biology", *Science Progress*, vol. 21, no. 82, pp. 341-343, 1925.

[17] V. Volterra, "Variations and Fluctuations of the Number of Individuals in Animal Species living together", *ICES Journal of Marine Science*, vol. 3, no. 1, pp. 3-51, 1926.

[18] U. M. Ascher and L. R. Petzold, *Computer methods for ordinary differential equations and differential-algebraic equations*, 1st. ed. Philadelphia, USA: SIAM, 1998.

[19] T. R. Malthus, *An essay on the principle of population*. Chapter 1, p.13 in Oxford World's Classics reprint, 1798.

[20] J. Peterson, "Global Population Projections through the 21st Century: A Scenario for This Issue", *Population, Resources and Environment*, vol. 13, no. 3, pp. 134-141, 1984.

[21] P. F. Verhulst, "Notice sur la loi que la population poursuit dans son accroissement", *Correspondance mathématique et physique*, vol. 10, pp. 113-121, 1838.

[22] S. H. Strogatz, *Nonlinear Dynamics and Chaos: with applications to Physics, Biology, Chemestry, and Engineering*, Chapter 2, 2nd. ed., Florida, USA: CRC Press, 2018.

[23] W. C. Allee and E. S. Bowen, "Studies in animal aggregations: Mass protection against colloidal silver among goldfishes", *Journal of Experimental Zoology*, vol. 61, no.2, pp. 185-207, 1932.

[24] D. Ludwig, D. D. Jones, and C. S. Holling, "Qualitative analysis of insect outbreak systems: the spruce budworm and forest", *J. Anim. Ecol.* vol. 47, no. 315, 1978.

[25] J. Al-Omari, "Stability and Optimal Harvesting in Lotka-Volterra Competition Model for TwoSpecies with Stage Structure", *Kyungpook Mathematical Journal* vol. 47, pp. 31-56, 2007.

[26] H. Cho, A. Lewis, K. Storey, and H. Byrne, "Designing experimental conditions to use the Lotka-Volterra model to infer tumor cell line interaction types ", *Journal of Theoretical Biology*, vol 559, 2023.

[27] J. D. Murray, *Mathematical Biology I: an introduction*, Chapter 3, 3rd. ed., Springer, 2002.

[28] M.A. Feitelson, A. Arzumanyan, R.J. Kulathinal, S.W. Blain, R.F. Holcombe, J. Mahajna, M. Marino, M.L Martinez-Chantar, R. Nawroth, I. Sanchez-Garcia, D. Sharma, N.K. Saxena, N. Singh, P.J. Vlachostergios, S. Guo, et al., "Sustained proliferation in cancer: Mechanisms and novel therapeutic targets", *Semin. Cancer Biol.*, pp. 25-54, 2015.

[29] H. Fort, *Ecological Modelling and Ecophysics: Agricultural and environmental applications*, Chapter 1 and Chapter 2, 1st. ed., Bristol, England: IOP Publishing, 2020.

[30] M. Kreft, M. Lukšič, T. M. Zorec, M. Prebil and R. Zorec, "Diffusion of D-glucose measured in the cytosol of a single astrocyte", *Cell Mol. Life Sci.*, vol. 70, no. 8, pp.1483-1492, 2013.

[31] P. Erdős, A. Rényi, "On Random Graphs", *Publicationes Mathematicae*, vol. 6, pp. 290–297, 1959.

[32] A. L. Barabási and R. Albert, "Emergence of scaling in random networks", *Science*, vol. 286, pp. 509-512, 1999.

[33] S. Bernard, H. Herzel, "Why do cells cycle with a 24 hour period?", *Genome Inform.*, vol. 17, no. 1, pp. 72-79, 2006.

[34] P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, "Generalized Louvain Method for Community Detection in Large Networks", *ISDA*, vol. 10, pp. 1109, 2011.