Insights into resource consumption, cross-feeding, system collapse, stability and biodiversity from an artificial ecosystem

Journal of the Royal Society Interface

Yu Liu*, David Sumpter

Department of Mathematics, Uppsala University, 75105 Uppsala, Sweden

E-mail: yu.liu@math.uu.se

S. Supplementary Material

(a) Energy conservation in the model

For the model itself, there might at first sight appear to be no constraints of energy, because metabolites can be easily transformed cyclicly and organisms can be reproduced through chemical reactions both in the forward and backward direction. For example, in the 3-metabolite system, species s_{11} reproduces via

$$s_{11} + \bar{1} + \bar{1} \to 2s_{11} + \bar{2}, \tag{S.1}$$

and species s_{22} reproduces via

$$s_{22} + \bar{2} + \bar{2} \to 2s_{22} + \bar{1}. \tag{S.2}$$

Thus, metabolite $\bar{1}$ can be transformed to $\bar{2}$ and $\bar{2}$ can be transformed to $\bar{1}$, and for both directions, organisms can be reproduced. It seems that the organisms are reproduced out of nothing.

However, since the transformation is based on modular addition, there are constraints automatically. That is, in the case above, two metabolites $\bar{1}$ can be transformed to one $\bar{2}$, but two $\bar{2}$ can only be transformed to one $\bar{1}$, rather than four or more $\bar{1}$. If we assume that $\bar{1}$ contains 2 units of energy and $\bar{2}$ contains 3 units of energy, energy conservation law will not be violated. In (S.1), two metabolites $\bar{1}$ contain 4 units of energy and one $\bar{2}$ contains 3 units of energy. The "missing" 1 unit of energy can be interpreted as being used to reproduce one organism s_{11}. Similarly, the missing 4 units of energy in (S.2) are used to reproduce one organism of s_{22}.

Generally, in n-metabolite system, if we assume that metabolite $\bar{i}$ contains $i + 1$ units of energy, it can always be interpreted in the way so that the energy conservation law holds. It is proved as follows. In n-metabolite system, all the chemical reactions for reproduction can be written as:

$$s_{ij} + \bar{i} + \bar{j} \to \begin{cases} 2s_{ij} + \overline{i+j}, & \text{if } i + j < n + 1 \\ 2s_{ij} + \bar{1}, & \text{if } i + j = n + 1 \\ 2s_{ij} + \bar{1} + \overline{(i+j) \bmod (n+1)}, & \text{if } i + j > n + 1 \end{cases}$$

As we assumed, $\bar{i}$ contains $i + 1$ units of energy, so the energy contained within the metabolites in the left hand side is $i + j + 2$. On the other hand, in the right hand side, for the first case, the energy contained within the metabolites is $i + j + 1$, so the energy used for reproduction is $(i + j + 2) - (i + j + 1) = 1 > 0$. For the second case, the energy used for reproduction is $(i + j + 2) - 2 = i + j = n + 1 > 0$. For the third case, the energy is $(i + j + 2) - (2 + (i + j - (n + 1) + 1)) = n > 0$. Therefore, for any case, there is some energy used to reproduction. Proof ends.

Note that the way to interpret those chemical reactions so that the conservation law of energy holds does not affect the dynamics of the system, but it makes the basis of the model more solid.

(b) Differential equations for the 3-metabolite system

We can write down the deterministic differential equations for the 3-metabolite system based on equation (2.6). The net production of metabolite $\bar{1}$ for one organism of species s_{11} is $A_{111} = -2$, and the net production of $\bar{1}$ for one of s_{12} is $A_{112} = -1$. Similarly, $A_{113} = 0$, $A_{122} = 1$, $A_{123} = 2$ and $A_{133} = 1$. As for the net production of metabolite $\bar{2}$, then $A_{211} = 1$, $A_{212} = -1$, $A_{213} = 0$, $A_{222} = -2$, $A_{223} = -1$ and $A_{233} = 1$. As for the net production of $\bar{3}$, then $A_{311} = 0$, $A_{312} = 1$, $A_{313} = -1$, $A_{322} = 0$, $A_{323} = -1$ and $A_{333} = -2$. Therefore, the equations in the 3-metabolite system can be written as, with inflow of the metabolite $\bar{2}$, namely $u = 2$:

$$\frac{dR_1}{dt} = -2S_{11}q(R_1, R_1) - S_{12}q(R_1, R_2) + S_{22}q(R_2, R_2) + 2S_{23}q(R_2, R_3) + S_{33}q(R_3, R_3)$$

$$\frac{dR_2}{dt} = \mu + S_{11}q(R_1, R_1) - S_{12}q(R_1, R_2) - 2S_{22}q(R_2, R_2) - S_{23}q(R_2, R_3) + S_{33}q(R_3, R_3)$$

$$\frac{dR_3}{dt} = S_{12}q(R_1, R_2) - S_{13}q(R_1, R_3) - S_{23}q(R_2, R_3) - 2S_{33}q(R_3, R_3)$$

$$\frac{dS_{11}}{dt} = S_{11}(2q(R_1, R_1) - 1)$$

$$\frac{dS_{12}}{dt} = S_{12}(2q(R_1, R_2) - 1)$$

$$\frac{dS_{13}}{dt} = S_{13}(2q(R_1, R_3) - 1)$$

$$\frac{dS_{22}}{dt} = S_{22}(2q(R_2, R_2) - 1)$$

$$\frac{dS_{23}}{dt} = S_{23}(2q(R_2, R_3) - 1)$$

$$\frac{dS_{33}}{dt} = S_{33}(2q(R_3, R_3) - 1)$$

(c) Visualisation of demographic structures

Species are represented by circles. The area of the circle is proportional to the size of the population.

If species A produces the metabolites species B needs, we say A supports B, represented by an arrow pointing from A to B. The shading of the arrow is proportional to the total number of metabolites that B obtains from A. The more B obtains, the darker the arrow is. The shading thus represents the actual support B gets from A. In this section, we always take the 2-metabolite system as an example, as Fig. 2. Species s_{11} ($\bar{1} + \bar{1} \to \bar{2}$) whose population is denoted by S_{11}, supports species s_{22} ($\bar{2} + \bar{2} \to \bar{1} + \bar{1}$) whose population is denoted by S_{22}. The number of metabolites $\bar{2}$ that those s_{11} could potentially provide is S_{11}, while that those s_{22} could potentially obtain is $2S_{22}$. So the total number of metabolites that s_{22} obtains from s_{11} is $\min(S_{11}, 2S_{22})$. Therefore, the shading of this arrow is proportional to $\min(S_{11}, 2S_{22})$. For some species, it produces the metabolite needed by itself, such as species s_{12} ($\bar{1} + \bar{2} \to \bar{1}$). We then draw an shaded arrow pointing to itself.

If two species A and B have a common need for certain metabolites, we say that they compete with each other, represented by a dash line connecting them. The shading of the dash line is proportional to the actual competition. For example, both species s_{11} and s_{12} need metabolite $\bar{1}$. The actual need of $\bar{1}$ by S_{11} is $2S_{11}$, while the actual need of $\bar{1}$ by S_{12} is S_{12}. So, they compete for $\min(2S_{11}, S_{12})$ of $\bar{1}$. Therefore, the shading of this dash line is proportional to $\min(2S_{11}, S_{12})$. The larger $\min(2S_{11}, S_{12})$ is, the severe the competition is, so the darker the dash line is.

The inflowing metabolite is illustrated by arrows pointing to species which needs the metabolite. For example, as in Fig. 2, species s_{11} and s_{12} need the inflowing metabolite $\bar{1}$. So there are two arrows pointing to s_{11} and s_{12}, respectively. The shading of the arrows is proportional to $2S_{11}$ and S_{12}, respectively. This scheme is equivalent to consider the inflow as an "imaginary species". The imaginary species produces the inflowing metabolite by taking in nothing, and its population is infinite.

(d) Classification of demographic structures from the data of simulations

We elaborate the scheme along with an example in the 3-metabolite system where $S_{11} = 350$, $S_{12} = 489$, $S_{13} = 12$, $S_{22} = 172$, $S_{23} = 510$ and $S_{33} = 8$.

Firstly, sort the species by population from large to small, and calculate the proportions, respectively. Then we have that s_{23} accounts for 33.1%, s_{12} accounts for 31.7% , s_{11} accounts for 22.7%, s_{22} accounts for 11.2%, s_{13} accounts for 0.8% and s_{33} accounts for 0.5%.

Secondly, add up the proportions of species one by one from large to small until the sum exceeds a predefined parameter Γ. For example, we set $\Gamma = 90\%$, so the sum of the proportions of the first four species s_{23}, s_{12}, s_{11} and s_{22} is 98.7%, having exceeded 90%. Then we say that this demographic structure consists of the four species. By this, we find the main species, and neglect others.

Thirdly, we divide 100% into ξ parts, so each species can be put into one of the ξ ranges. Then we say that the species fallen in the same range account for the same proportion. For example, we set $\xi = 20$, so species s_{23} and s_{12} are in the range between 30% and 35%, species s_{11} is in the range between 20% and 25%, and species s_{22} is in the range between 10% and 15%.

Then we say that this demographic structure consists of 22.5% s_{11}, 32.5% s_{12}, 0% s_{13}, 12.5% s_{22}, 32.5% s_{23} and 0% s_{33}. For all the demographic structures discussed in this paper, we set $\Gamma = 90\%$ and $\xi = 20$.

In some cases, it is more convenient to only consider what species constitute the whole population, with no need to mention the particular proportions. For example, for the demographic structure represented by the red line in Fig. 3, we just say that it is composed of the three species s_{11}, s_{12} and s_{22}, without mentioning their proportions respectively. So that means, this demographic structure could refer to many demographic structures consisting of the three species but with different proportions.

(e) Analytical analysis for the system $n = 2$ and $u = 1$

The differential equations for the system with $n = 2$ and $u = 1$ are written as

$$\frac{dR_1}{dt} = \mu - 2S_{11}q(R_1, R_1) + 2S_{22}q(R_2, R_2)$$

$$\frac{dR_2}{dt} = S_{11}q(R_1, R_1) - S_{12}q(R_1, R_2) - 2S_{22}q(R_2, R_2)$$

$$\frac{dS_{11}}{dt} = S_{11}(2q(R_1, R_1) - 1)$$

$$\frac{dS_{12}}{dt} = S_{12}(2q(R_1, R_2) - 1)$$

$$\frac{dS_{22}}{dt} = S_{22}(2q(R_2, R_2) - 1)$$

There are two steady states in total:

$$R_1 = (2a + 1 + \sqrt{8a^2 + 1})/2, \quad R_2 = (2a + 1 + \sqrt{8a^2 + 1})/2, \quad S_{11} = 2\mu, \quad S_{12} = 0, \quad S_{22} = \mu;$$

$$R_1 = (2a + 1 + \sqrt{8a^2 + 1})/2, \quad R_2 = (2a - 1 + \sqrt{8a^2 + 1})/2, \quad S_{11} = \mu, \quad S_{12} = \mu, \quad S_{22} = 0.$$

As $a = 5$, so $(2a + 1 + \sqrt{8a^2 + 1})/2 \approx 12.59$ and $(2a - 1 + \sqrt{8a^2 + 1})/2 \approx 11.59$.

Then check the stabilities of the two steady states by linearising the equations near each steady state. For the first steady state, there are five eigenvalues for the corresponding Jacobian, namely -62.59, -8.14, -1.14, -1.02 and 0.02 respectively. There is one positive eigenvalue, so it is an unstable steady state (saddle point). For the second steady state, the five eigenvalues are -23.25, -5.27, -1.04, -1.23 and -0.05 which are all negative, so it is a stable steady state.

(f) Every steady state consists of species loops, but not vice versa

First, we show that every steady state consists of species loops. Take the differential equations corresponding to the system $n = 3$ and $u = 2$ as an example (see Section (b) of the supplementary material). To get steady states, the left hand side of all the nine differential equations should be zero. So for the first three equations, we have

$$0 = -2S_{11}q(R_1, R_1) - S_{12}q(R_1, R_2) + S_{22}q(R_2, R_2) + 2S_{23}q(R_2, R_3) + S_{33}q(R_3, R_3)$$

$$0 = \mu + S_{11}q(R_1, R_1) - S_{12}q(R_1, R_2) - 2S_{22}q(R_2, R_2) - S_{23}q(R_2, R_3) + S_{33}q(R_3, R_3) \quad \text{(S.3)}$$

$$0 = S_{12}q(R_1, R_2) - S_{13}q(R_1, R_3) - S_{23}q(R_2, R_3) - 2S_{33}q(R_3, R_3)$$

and for the last six equations, we have

$$\begin{cases} S_{ij} = 0 \quad \text{and} \quad q(R_i, R_j) \text{ is any real number}, \quad \text{or} \\ S_{ij} \neq 0 \quad \text{and} \quad q(R_i, R_j) = 1/2 \end{cases} \quad \text{(S.4)}$$

where $i, j = 1, 2, 3$ and $j \geqslant i$. Note that the constrains $q(R_i, R_j) = 1/2$ are used to solve R_1, R_2 and R_3, and they constrain S_{ij} only in the way that, e.g., it is impossible that S_{11}, S_{12} and S_{22} are simultaneously nonzero, since it requires $q(R_1, R_1) = q(R_1, R_2) = q(R_2, R_2) = 1/2$ which turns out to have no solution.

Combining (S.3) and (S.4), and denoting $2S_{ij}q(R_i, R_j)$ as h_{ij} for convenience, we get

$$0 = -2h_{11} - h_{12} + h_{22} + 2h_{23} + h_{33} \quad \text{(S.5)}$$

$$-2\mu = h_{11} - h_{12} - 2h_{22} - h_{23} + h_{33} \quad \text{(S.6)}$$

$$0 = h_{12} - h_{13} - h_{23} - 2h_{33} \quad \text{(S.7)}$$

where h_{ij} is either zero, then no other constrain for the corresponding $q(R_i, R_j)$, or nonzero (that is, $h_{ij} = S_{ij}$), then the corresponding $q(R_i, R_j)$ should be $1/2$.

By letting h_{ij} be zero or nonzero, we can solve the equation system (S.5) - (S.7). There must exist nontrivial solutions, because the number of independent equations is smaller than the number of variables. Note the fact that, in the original differential equations, the right hand side represents the net production rate of each metabolite. Therefore, (S.5) and (S.7) mean that the net production rate of metabolite $\bar{1}$ and $\bar{3}$ is zero, and (S.6) means that for metabolite $\bar{2}$ the inflow rate is equal to the consumption rate. So we have shown that any steady state of the original differential equations satisfies that all the inflow is fully consumed and no other metabolites are produced. That is to say, every steady state consists of species loops.

Second, by a counterexample we show that not every species loop corresponds to a steady state:

$$\begin{cases} h_{11} = & 2\mu/3 \\ h_{12} = & 4\mu/3 - 1 \\ h_{13} = & 0 \\ h_{22} = & 1 \\ h_{23} = & 4\mu/3 - 1 \\ h_{33} = & 0 \end{cases} \quad \text{(S.8)}$$

is a solution of the equation system (S.5) - (S.7). Note that $h_{ij} = S_{ij}$ if $h_{ij} \neq 0$. So this solution represents species loops. To make it clearer, take $\mu = 3$, and then we have $S_{11} = 2$, $S_{12} = 3$, $S_{22} = 1$,

$S_{23} = 3$, $S_{13} = 0$ and $S_{33} = 0$. One can check that it is a species loop indeed. However, according to (S.4), since h_{11}, h_{12} and h_{22} are nonzero, we must have $q(R_1, R_1) = q(R_1, R_2) = q(R_2, R_2) = 1/2$ which has no solutions. So (S.8) satisfies (S.3) but not (S.4). Therefore, the species loop represented by (S.8) is not a steady state of the corresponding differential equations.

For other systems with different n and u, the same argument applies.

(g) Number of species loops in equilibrium

If in equilibrium the whole population is composed of one species loop, the number of this species loop always fluctuates around $2\mu/\Omega$. It is a general property of our artificial ecosystem.

Assume that in an equilibrium the whole population is composed of one specific species loop with Ω. Let m be the total number of this species loop in equilibrium at any time step. After one unit time of the simulation, namely one generation, on average every organism has been selected and completed a life cycle. Let r be the number of the species loops that have taken in required metabolites during this generation. So after this generation, there are $2r$ species loops because organisms which take in required metabolites will reproduce. Since it is in equilibrium, the number of species loops should be kept the same, so

$$m = 2r.$$

On the other hand, the number of consumed inflowing metabolites is $\Omega \cdot r$. It should be equal to the number of metabolites inflowing into the system during one generation, that is,

$$\Omega \cdot r = \mu.$$

The two equations give $m = 2\mu/\Omega$ where m is the number of species loops in equilibrium. Since the simulation is stochastic, one will find the number of species loops in equilibrium fluctuates around $2\mu/\Omega$. The derivation above has nothing to do with the number of types of metabolites n and the inflowing metabolite u, so the derivation is general.

(h) Calculating the proportions of different species loops

As we discussed, in the artificial ecosystem, at any one time various species loops persist, self-organizing to consume up all the available resources. That is, any demographic structure which is able to persist in the ecosystem can be considered as being constituted by different species loops. This is a general property for any n-metabolite system with any inflow.

When n is small, it is easy to work out the proportions of different species loops given a certain demographic structure, but when n is large, it is far from straightforward. Here the scheme for calculating the proportions for small n-metabolite system is first described, and the scheme for arbitrary n-metabolite system is elaborated afterwards.

For the small n-metabolite system, we take the demographic structure in Fig. 2(c) ($n = 2$) as an example, where $S_{11} = 642$, $S_{12} = 355$ and $S_{22} = 133$. By observation, we guess it is composed of two species loops, $s_{11} + s_{12}$ and $2s_{11} + s_{22}$. Then assume the number of the former species loop is x, and the number of the latter one is y. Then we should have the following equations:

$$
\begin{aligned}
x + 2y &\approx 642 \\
x &\approx 355 \\
y &\approx 133
\end{aligned}
$$

Since the simulation is stochastic, we have already allowed errors in these equations.

The three equations correspond to three straight lines in the coordinate of x and y. The true values of x and y correspond to a point (x, y) such that the sum of the distances between itself

and the three lines is minimal. That is, the following expression has the minimal value

$$((x + 2y - 642)/\sqrt{5})^2 + (x - 355)^2 + (y - 133)^2$$

When $x = 357.1$ and $y = 137.2$, the above expression has the minimal value. Therefore, approximately this demographic structure consists of $357.1/(357.1 + 137.2) \approx 72.2\%$ of the species loop $s_{11} + s_{12}$ and 27.8% of the species loop $2s_{11} + s_{22}$.

Now for the arbitrary n-metabolite system, I elaborate the general scheme, taking a demographic structure appeared in an evolution in the 6-metabolite system as an example, as shown in Fig. S1.
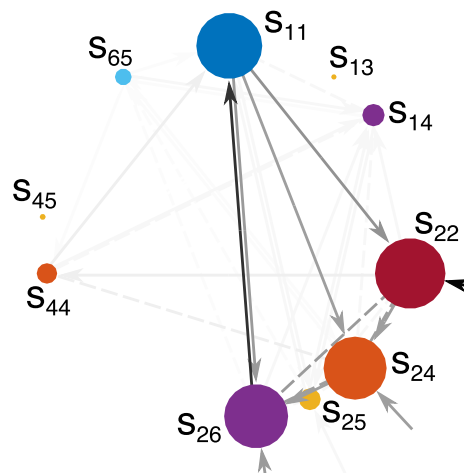

Figure S1. The demographic structure arising in an evolution in the 6-metabolite system with inflowing metabolite $\bar{2}$, where $S_{11} = 1221, S_{13} = 2, S_{14} = 116, S_{22} = 1403, S_{24} = 1112, S_{25} = 110, S_{26} = 1147, S_{44} = 95, S_{45} = 3, S_{65} = 20$ and other species are zeros (21 species in total).

The demographic structure is expressed by a vector recording the populations of all the non-zero species. In this scheme, there are two main procedures. First, proportionally reduce the population vector such that the "reduced population vector" is a superposition of a certain set of species loops, and its elements are integers and as small as possible. The difficulty here is how to cancel out the noise of the demographic structure. Second, find the certain set of species loops constituting the reduced population vector, and work out their proportions respectively.

The steps of the first procedure are as follows:

(i) Create a vector $\mathbf{p}$ of which the element p_i is the population of non-zero species (neglecting species with extremely low population).
In the example, $\mathbf{p} = [1221, 116, 1403, 1112, 110, 1147, 95]^{\mathsf{T}}$ (the superscript T means the transpose of a vector or matrix) successively denoting the population of species s_{11}, s_{14}, s_{22}, s_{24}, s_{25}, s_{26} and s_{44} (the extremely low population species s_{13}, s_{45} and s_{65} are neglected).

(ii) Create a matrix $\mathbf{A}$ of which the entry A_{ki} represents the net production of the kth metabolite for one organism of the ith species in the population vector, that is, s_{11} is the 1st species, s_{14} is the 2nd species, s_{22} is the 3rd species and so on.

In the example,

$$\mathbf{A} = \begin{bmatrix} -2 & -1 & 0 & 0 & 1 & 2 & 2 \\ 1 & 0 & -2 & -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & -1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 \end{bmatrix}$$

For (e.g.) the 1st row, the 1st element -2 means one organism of species s_{11} produces -2 of $\bar{1}$ since it takes in 2 of $\bar{1}$; the 2nd element -1 means one of s_{14} produces -1 of $\bar{1}$; and the 7th element 2 means one s_{44} produces 2 of $\bar{1}$. For the 4th row, the 4th element -1 means one s_{24} produces -1 of $\bar{4}$; and the 5th element 0 means one s_{25} produces 0 of $\bar{4}$; and so on.

Delete the row corresponding to the inflowing metabolite, and let $\mathbf{A}\mathbf{x} = \mathbf{0}$ (where $\mathbf{x}$ is the reduced population vector), meaning that the net production of every kind of metabolite (except for the inflowing metabolite) is zero. Also delete the zero row since it does not provide any useful information. Then we have

$$\mathbf{A} = \begin{bmatrix} -2 & -1 & 0 & 0 & 1 & 2 & 2 \\ 0 & -1 & 1 & -1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 \end{bmatrix}$$

and

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} -2 & -1 & 0 & 0 & 1 & 2 & 2 \\ 0 & -1 & 1 & -1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where x_1, x_2, x_3, x_4, x_5, x_6 and x_7 are the reduced population of species s_{11}, s_{14}, s_{22}, s_{24}, s_{25}, s_{26} and s_{44}, respectively.

(iii) Get the "rational" basis $\mathbf{V}$ for the null space of $\mathbf{A}$ obtained from the reduced row echelon form (it is a routine for linear algebra, referring to the MATLAB function *null*). Then the reduced population vector $\mathbf{x}$ can be written as the linear combination of the basis with m free parameters where m is the number of columns of $\mathbf{V}$.

In this example,

$$\mathbf{V} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 2 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ denoted as } [\mathbf{v}_1|\mathbf{v}_2|\mathbf{v}_3]$$

Note that the submatrix of $\mathbf{V}$ from 5th to 7th row is in echelon form (that is what I mean by "obtained from the reduced row echelon form"), and that in most cases the entries in the submatrix from the 1st to 4th row are rational numbers (here they happen to be integers). The number of columns of $\mathbf{V}$ is three, so there are three free parameters. Then

we have

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \alpha \mathbf{v}_1 + \beta \mathbf{v}_2 + \gamma \mathbf{v}_3 = \alpha \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \gamma \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{S.9}$$

where α, β and γ are any real numbers.

(iv) Choose a range for all the free parameters. Substitute each value within these ranges into $\mathbf{x}$. Record all the valued $\mathbf{x}$ of which all the elements are non-negative integers, denoted as $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n$. So every $\mathbf{y}_i$ is thus a group of organisms which as a whole only consume the inflowing metabolites but produce nothing. That is to say, $\mathbf{y}_i$ could be a species loop, but not necessarily since it may not be a minimum set of organisms (referring to the definition of species loop).

From (S.9) we see that in order for every element of $\mathbf{x}$ to be non-negative integer, all the free parameters α, β and γ have to be non-negative integers. In the example, I choose α, β and γ to be any value in the set $\{0, 1, 2, \cdots, 50\}$. And then I get 132651 possible $\mathbf{y}$.

(v) For each possible $\mathbf{y}_i$, one can get a vector $\mathbf{q}_i$ whose elements are quotients of $\mathbf{p}$ dividing $\mathbf{y}_i$ by element. For example, for $\mathbf{y}_1$, the 1st element in $\mathbf{q}_1$ is the quotient dividing p_1 by the 1st element of $\mathbf{y}_1$, the 2nd element in $\mathbf{q}_1$ is the quotient dividing p_2 by the 2nd element of $\mathbf{y}_1$, and so on. Then for each vector $\mathbf{q}_i$, calculate the average, denoted as ϕ_i, and the standard deviation of the elements, denoted as δ_i. In this example, obviously I get the average and standard deviation of the 132651 $\mathbf{q}$.

(vi) Sort all of $\mathbf{y}$ in ascending order according to its corresponding δ, denoted as $\mathbf{y_s}$. Then the first one, $\mathbf{y_{s}}_1$, is the best candidate for the reduced population vector, which is deviated the least from the original population vector $\mathbf{p}$. If its corresponding δ is small enough (namely $\delta/\phi \ll 1$), let it be the reduced population vector, denoted as $\mathbf{b}$; otherwise, repeat from step (iv) with choosing larger ranges until the sorted first δ is small enough. The reason why sometimes the sorted first δ is not small is that the ranges one chooses for the free parameters (here are α, β and γ) are too small. However, larger range requires more computations which increase quickly with problem size. So there is a tradeoff.

In this example, ϕ and δ of $\mathbf{y_{s}}_1$ are 22.96 and 0.58, respectively ($0.58/22.96 \approx 0.0025 \ll 1$), so I take it as the reduced population vector $\mathbf{b}$ which is $[53, 5, 62, 49, 5, 49, 4]^\mathsf{T}$.

Up until here, we get the reduced population vector $\mathbf{b}$ and thus finish the first procedure. The second main procedure is to find which species loops constitute $\mathbf{b}$. The steps are as follows:

(vii) For each $\mathbf{y_{s}}_i$, check whether there is at least one element which is larger than the corresponding element in $\mathbf{b}$. If yes, throw it away, and put all the other $\mathbf{y_s}$ (change notation to $\mathbf{z}$) into a matrix $\mathbf{Z} = [\mathbf{z}_1 | \mathbf{z}_2 | \cdots | \mathbf{z}_n]$.

For example, this $\mathbf{y_{s}}_i = [53, 5, 63, 48, 5, 48, 5]^\mathsf{T}$ will be thrown away because the 3rd and 7th element 63 and 5 are larger than the corresponding element in $\mathbf{b}$ (62 and 4). After all, in this example, I get the matrix $\mathbf{Z}$ which only has 1498 columns, out of the original 132651 columns.

(viii) If all the elements of $\mathbf{z}_j$ are equal to or smaller than the corresponding element of $\mathbf{z}_i$, we say $\mathbf{z}_j$ is non-larger than $\mathbf{z}_i$. Count the number of non-larger vectors in $\mathbf{Z}$ for every $\mathbf{z}_i$, denoted as h_i. And then sort $\mathbf{z}$ in ascending order according to h, denoted as $\mathbf{z_s}$ and $\mathbf{Z}$ as $\mathbf{Z_s}$.

(ix) Since the number of linearly independent columns of $\mathbf{Z_s}$ must be m, randomly select m of $\mathbf{z_{s}}_i$, and put them together as a matrix $\mathbf{R}$. If (1) all the h's of these $\mathbf{z_{s}}_i$ are zero, (2) these $\mathbf{z_{s}}_i$ are linearly independent, and (3) all the elements in the vector $\mathbf{R}^{-1}\mathbf{b}$ are non-negative integers, record the m of $\mathbf{z_{s}}_i$ and end the procedure; otherwise, repeat from

randomly selecting m of $\mathbf{z}_{\mathbf{s}i}$. Note that the computation should be much faster if one first tries $\mathbf{z}_{\mathbf{s}i}$ in the left end of $\mathbf{Z}_{\mathbf{s}}$, since they are more likely to have $h = 0$ (that is why we sort $\mathbf{Z}$ to $\mathbf{Z}_{\mathbf{s}}$ according to h).

Then these m of $\mathbf{z}_{\mathbf{s}i}$ are the species loops constituting the demographic structure, and the elements in the vector $\mathbf{R}^{-1}\mathbf{b}$ are the corresponding portions of each species loop.

In this example, we get the answer that the reduced population vector $\mathbf{b} = [53, 5, 62, 49, 5, 49, 4]^{\mathsf{T}}$ consists of 5 of $[0, 1, 1, 0, 1, 0, 0]^{\mathsf{T}}$, 4 of $[1, 0, 2, 0, 0, 0, 1]^{\mathsf{T}}$ and 49 of $[1, 0, 1, 1, 0, 1, 0]^{\mathsf{T}}$. That is, this demographic structure shown in Fig. S1 consists of 5 portions of the species loop $s_{14} + s_{22} + s_{25}$ ($\Omega = 3$), 4 portions of the species loop $s_{11} + 2s_{22} + s_{44}$ ($\Omega = 3$), and 49 portions of the species loop $s_{11} + s_{22} + s_{24} + s_{26}$ ($\Omega = 3$) (the proportions are consequently 8.6%, 6.9% and 84.5%, respectively).

Finally, note that choosing different ranges for the m free parameters (namely α, β and γ here), one will get different reduced population vector $\mathbf{b}$. But as long as the standard deviation δ for the first $\mathbf{y}_{\mathbf{s}}$ is small enough, the final results converge, that is, the compositions of species loops and their proportions are almost the same. Also note that, this algorithm does find these species loops and their corresponding proportions, but as the system in question gets larger and larger, it takes longer and longer computational time. For a normal PC, it takes several days to get the answer for the 6-metabolite system.

(i) Analytical analysis for the system $n = 3$ and $u = 2$

Refer to the differential equations for the system with $n = 3$ and $u = 2$ shown in Section (b) of the supplementary material. We found the following steady states corresponding to the demographic structures in Fig. 5, although there are more steady states:

$$R_1 = \beta, \quad R_2 = \beta, \quad R_3 = \psi, \quad S_{11} = \frac{2}{3}\mu, \quad S_{12} = 0,$$

$$S_{13} = 0, \quad S_{22} = \frac{4}{3}\mu, \quad S_{23} = 0, \quad S_{33} = 0, \quad (\psi \in \mathbf{R}^+);$$

$$R_1 = \sigma, \quad R_2 = \beta, \quad R_3 = \beta, \quad S_{11} = 0, \quad S_{12} = \frac{4}{3}\mu,$$

$$S_{13} = 0, \quad S_{22} = \frac{2}{3}\mu, \quad S_{23} = 0, \quad S_{33} = \frac{2}{3}\mu;$$

$$R_1 = \beta, \quad R_2 = \sigma, \quad R_3 = \beta, \quad S_{11} = \frac{2}{3}\mu, \quad S_{12} = \frac{4}{3}\mu,$$

$$S_{13} = 0, \quad S_{22} = 0, \quad S_{23} = \frac{4}{3}\mu, \quad S_{33} = 0.$$

where $\beta \equiv (2a + 1 + \sqrt{8a^2 + 1})/2$ and $\sigma \equiv (2a - 1 + \sqrt{8a^2 + 1})/2$. As $a = 5$, so $\beta \approx 12.59$ and $\sigma \approx 11.59$. The first set of steady states corresponds to the demographic structure in Fig. 5(a), the second steady state corresponds to the one in Fig. 5(b), and the third steady state corresponds to the one in Fig. 5(d). But there is no steady state corresponding to the one in Fig. 5(c).

For the stability of the first set of steady states, the nine eigenvalues of the Jacobian with varying ψ are shown in Fig. S2(a). We see that many eigenvalues are kept the same when ψ is large enough and two of them are always very negative (-37.29 and -9.15 respectively), so we zoom in to the interesting part to see the other seven eigenvalues, as shown in Fig. S2(b). There are four constant eigenvalues, two negative (-1.12 and -1.03 respectively), one 0, and one 0.02. There are two eigenvalues always the same, and they change from negative to positive at $\psi = \sigma$. There is another eigenvalue changing from negative to positive at $\psi = \beta$. All in all, this set of steady states are unstable. Particularly, there are one positive eigenvalue when $\psi < \sigma$ and four positive eigenvalues when $\psi > \beta$, which means that when ψ is larger than β, this set of steady states becomes much more unstable.
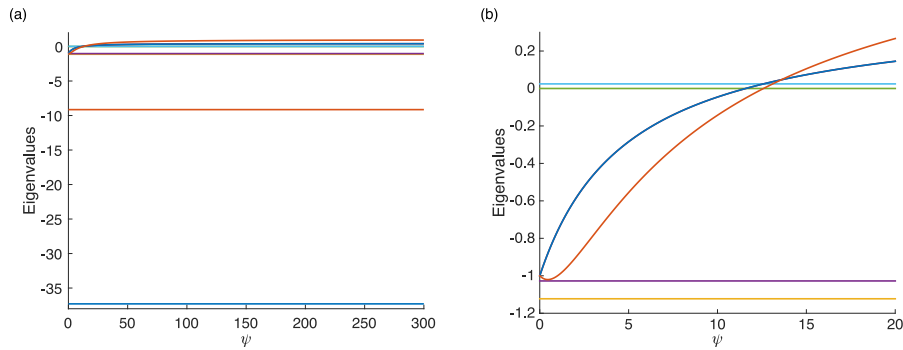
(a)

(b)

Figure S2. Eigenvalues of the Jacobian for the set of steady states with varying ψ. (a) All the nine eigenvalues with ψ varying from 0 to 300. (b) The enlarged figure of (a) around the interesting part, namely varying ψ between 0 and 20. Here only seven eigenvalues are shown, and the other two very negative constant eigenvalues -37.29 and -9.15 are not shown.

The nine eigenvalues of the corresponding Jacobian for the second steady state are -29.71, -12.75, -2.11, -1.90, -1.09, -1.03, -0.05, 0 and 0.02, respectively. Besides the seven negative eigenvalues, there are one zero and one positive eigenvalue. Thus, it is an unstable steady state.

For the stability of the third steady state, the nine eigenvalues of the corresponding Jacobian are -29.57, -13.30, $-1.80 \pm 0.60i$, -1.08, -1.03, -0.05, 0 and 0.02, respectively. Besides the seven negative eigenvalues (among them, there are two complex eigenvalues whose real parts are negative), there are one zero and one positive eigenvalue. Thus, it is an unstable steady state.