

Polyhedral geometry and combinatorics of an autocatalytic ecosystem

Praful Gagrani¹, Victor Blanco^{2,3}, Eric Smith^{4,5,6,7}, David Baum^{1,8}

¹ *Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI 53715, USA*

² *Institute of Mathematics (IMAG), Universidad de Granada, Spain*

³ *Department of Quant. Methods for Economics & Business, Universidad de Granada, Spain*

⁴ *Earth-Life Science Institute, Tokyo Institute of Technology, Tokyo 152-8550, Japan*

⁵ *Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*

⁶ *Ronin Institute, 127 Haddon Place, Montclair, NJ 07043, USA*

⁷ *Department of Biology, Georgia Institute of Technology, Atlanta, GA 30332, USA*

⁸ *Department of Botany, University of Wisconsin-Madison, Madison, WI 53706, USA*

(Dated: November 14, 2023)

Developing a mathematical understanding of autocatalysis in reaction networks has both theoretical and practical implications. We review definitions of autocatalytic networks and prove some properties for minimal autocatalytic subnetworks (MASs). We show that it is possible to classify MASs in equivalence classes, and develop mathematical results about their behavior. We also provide linear-programming algorithms to exhaustively enumerate them and a scheme to visualize their polyhedral geometry and combinatorics. We then define *cluster chemical reaction networks*, a framework for coarse-graining real chemical reactions with positive integer conservation laws. We find that the size of the list of minimal autocatalytic subnetworks in a maximally connected cluster chemical reaction network with one conservation law grows exponentially in the number of species. We end our discussion with open questions concerning an ecosystem of autocatalytic subnetworks and multidisciplinary opportunities for future investigation.

Contents		4. Computational challenges in scaling	26
		C. Rule generated CCRNs	26
I. Introduction	1	VI. Discussion and future research	27
II. Chemical reaction networks and autocatalytic subnetworks	3	Acknowledgements	27
A. Preliminaries	3	Declarations	28
B. Chemical reaction networks	4	A. Deficiency theory and multistability in CCRNs	28
C. Input-output matrix pair and the stoichiometric matrix	4	B. Kernel of the stoichiometric matrix for the complete 1-constituent CCRN	29
D. Hypergraph flows and conservation laws	5	C. Deficiency-one algorithm for L=3 1-constituent CCRN	30
E. Formal, exclusive and stoichiometric autocatalysis	5	D. Details of the algorithm for detecting stoichiometrically autocatalytic motifs	30
1. Formal autocatalysis	6	1. Checking intersection of cones	32
2. Exclusive autocatalysis	7	E. Minimal Autocatalytic Subnetworks for CCRN with $L = 4$ and $L = 5$	33
3. Stoichiometric autocatalysis	8	References	34
III. Minimal autocatalytic subnetworks	8		
A. Properties	8		
B. Food-waste-member-core partition	11		
C. Flow, species and partition productive cones	12		
IV. Organizing an autocatalytic ecosystem	13		
A. Mathematical results	14		
B. Algorithms	16		
C. Geometry and visualization	17		
V. Cluster chemical reaction networks (CCRN) framework	19	I. INTRODUCTION	
A. Formalism	21	Chemical reaction network (CRN) theory offers a versatile mathematical framework in which to model complex systems, ranging from biochemistry and game theory [1], to the origins of life [2]. The usefulness of CRNs in modelling these phenomena stems from its ability to exhibit a wide range of nonlinear dynamics and it is widely	
B. Complete CCRN	23		
1. 1-constituent CCRN	23		
2. L=3 1-constituent CCRN	23		
3. Number theoretic result for 1-constituent CCRN of order two	24		

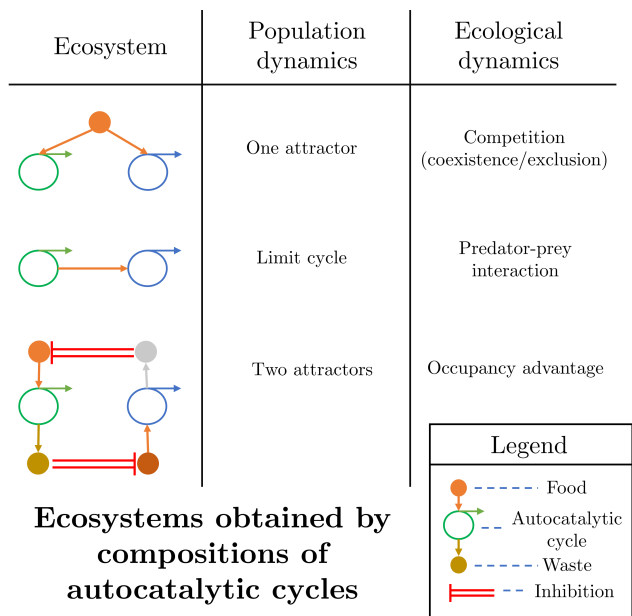


Figure 1: Ecosystems obtained by composing autocatalytic cycles. Each autocatalytic cycle consumes food and produce waste species. Two cycles can be composed in several ways, for example by sharing food (line one), the member species of one could be the food of other (line two), or the waste of one can deplete the food of another (line three). For a detailed account of such interactions and their ecological counterparts, see [7].

recognized that *autocatalysis* can be seen as a basis for many of them [3, 4]. Broadly, autocatalysis is framed as the ability of a given chemical species to make more copies of itself or otherwise promote its own production. Thus, from the perspective of kinetics, one would expect autocatalysts to be able to show super-linear growth, which is indeed possible [5, 6]. For some examples of nonlinear kinetics obtained by composing two autocatalytic cycles and their relevance to ecological dynamics, see Fig. 1.

A comprehensive understanding of autocatalysis in CRNs has great practical values, but its full mathematical treatment remains to be developed. In [6], Anderson et al. conjectured that it is impossible for a CRN to show a temporary speed-up of the reaction before settling down to reach an equilibrium if, at least, it is not *formally* autocatalytic. Recently, in [8], Vassena et al. show that this is not the case and there can be systems that show superlinear growth that do not exhibit autocatalysis. Also, it is suggested in [9] that understanding the interactions between autocatalytic and drainable subnetworks might provide a basis for obtaining a proof for the long-standing *persistence conjecture* [10].

Andersen et al. also defined a notion of *exclusive* autocatalysis for one species, which we generalize for multiple species. In [11], Blokhuis et al. defined a notion of auto-

catalysis, more restrictive than exclusive autocatalysis, and showed that, within reversible and non-ambiguous CRNs, it has only five types of minimal forms. They also found that autocatalysis is more abundant than previously thought, and simple reaction networks with a few reactions can contain very many cores in them.

In this work, we focus our attention on the topological notion of autocatalysis, one that can be ascertained purely from the reactions in a CRN. A CRN is autocatalytic in a set of core (autocatalytic) species if there exists a flow on the network such that all core species are necessarily being consumed while being produced at higher rate. We term the subnetwork that consists of all the reactions that have a strictly positive contribution to such a flow as an autocatalytic subnetwork. By imposing further constraints on the subnetworks, we resolve three nested categories of autocatalysis: formal, exclusive, and stoichiometric autocatalysis. We show that stoichiometric autocatalysis is similar to the autocatalysis of Blokhuis et al. (see Remark 11), and rederive some of their results for the broader class of exclusive autocatalysis. In particular, we prove that minimal exclusively autocatalytic subnetworks have the property that along with all core species being produced at a higher rate than they are consumed, at least one core species is produced and consumed in each reaction of the subnetwork. We then investigate the polyhedral geometry of these minimal autocatalytic subnetworks (MASs), provide algorithms for exhaustively enumerating them, and propose visualization schemes for understanding their combinatorics.

We employ the techniques developed in this work on a framework, which we term the *Cluster CRN* (CCRN). A brief description of the motivation and formalism of the CCRN framework is as follows. Molecules are composed of atoms and atomic conservation is a fundamental law of chemistry. While molecular structure plays a critical role in determining what atomic interchanges are possible in particular environments (and their rate constants), focusing on just atomic composition provides a simple abstraction that, at some level, must resemble real chemistry. In the most basic version with a single conserved quantity (such as a monomer or an atom), the list of all potential reactions is dictated only by arithmetic. By adding other dimensions it is theoretically possible to represent a lot of real chemistry. Adding dimensions can allow for polymers composed of multiple monomer types or for molecular formulae composed of more than one element, where the *dimension* is the number of monomer types or elements, respectively. Other dimensions can also be used to represent structural variation within a molecular formula or spatial variation, although doing so requires adding constraints besides arithmetic on allowed reaction rules. We thus posit that the CCRN framework is a very flexible coarse-graining of chemistry. Formally, a CCRN is a CRN with positive integer valued conservation laws. The investigations undertaken in this work show that even the simplest version CCRN (in one di-

mension) can reveal important principles regarding the abundance of MASs and their interactions.

The layout of the paper is as follows. In Sec. II, we briefly review chemical reaction networks (CRNs) and present a hierarchy of autocatalysis in their graph-theoretic and linear-algebraic manifestations. In Sec. III, we define MASs, prove their properties, organize them in equivalence classes, and explore some geometric quantities that are canonically defined for them. In Sec. IV, we consider CRNs with multiple MASs, which we term as an *autocatalytic ecosystem*. Given such a CRN, we provide a linear-programming algorithm to exhaustively enumerate the MASs and introduce a visualization scheme in which to represent their polyhedral geometry their combinatorics. In Sec. V, we present the cluster chemical reaction network (CCRN) framework and argue that it provides a natural coarse-graining of realistic chemical and biochemical reaction networks. We then explore the combinatorics of MASs within the CCRN framework and provide a worked example to demonstrate the geometry of autocatalysis in the stoichiometric subspace of species concentration. Using elementary number theoretic considerations, we also prove a combinatorial result about the list of MASs. Multiple examples from the CCRN framework are also included as examples at several points in this work. We conclude, in Sec. VI, with an overview of our contributions and possible avenues for future investigation.

II. CHEMICAL REACTION NETWORKS AND AUTOCATALYTIC SUBNETWORKS

The mathematical theory of chemical reaction networks (CRNs), pioneered by Horn, Jackson, and Feinberg [12] has ubiquitous applications in understanding nature [13, 14]. In particular, for diverse biological and chemical applications, the formalism allows for a notion of self-replication or *autocatalysis* where certain species cause an increase in the population of these same species. Consequently, there is value in rigorously understanding the organization of CRNs containing a collection of autocatalytic subnetworks, which we term as an *autocatalytic ecosystem*.

The concept of autocatalysis, dating back to the late 1800s [15, 16], has several distinct formulations in CRN theory. In this section, we explain the correspondence between graph-theoretic and linear-algebraic formulations of CRNs, and employ it to present a nested hierarchy of three types of autocatalysis: formal, exclusive, and stoichiometric. In particular, we provide conditions on a subset of reactions (subnetwork) of a CRN for them to be termed autocatalytic for some subset of autocatalytic species. Note that, for a subnetwork to be autocatalytic, we require that all the reactions participate with a strictly positive flux. So, while a CRN can be said to be autocatalytic if it contains an autocatalytic subnetwork, in our formulation, a subnetwork might lose its autocat-

alytic property if more reactions are added. This choice of definitions is made to simplify analysis (see Remark 11) and leads the way to an investigation of minimal autocatalytic subnetworks in the next section.

A. Preliminaries

Notation II.1. $\mathbf{0}$ denotes a vector of zeros.

Notation II.2. For a vector $\mathbf{v} = [v_1, v_2, \dots, v_n]^T$ in $\mathbb{R}^n$:

1. We call $\mathbf{v}$ positive and write $\mathbf{v} \gg \mathbf{0}$ if $v_i > 0$ for all $i \in [1, n]$.
2. We call $\mathbf{v}$ non-negative and write $\mathbf{v} \geq \mathbf{0}$ if $v_i \geq 0$ for all $i \in [1, n]$.
3. We call $\mathbf{v}$ semi-positive and write $\mathbf{v} > \mathbf{0}$ if $v_i \geq 0$ and $\mathbf{v} \neq \mathbf{0}$.

Notation II.3. For vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$, $\mathbf{v} \gg \mathbf{w}$, $\mathbf{v} \geq \mathbf{w}$, and $\mathbf{v} > \mathbf{w}$ mean $\mathbf{v} - \mathbf{w} \gg \mathbf{0}$, $\mathbf{v} - \mathbf{w} \geq \mathbf{0}$, and $\mathbf{v} - \mathbf{w} > \mathbf{0}$, respectively.

Definition II.1. For all vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k \in \mathbb{R}^n$:

1. Their *convex polyhedral cone* is denoted by

$$\text{Cone}(\{\mathbf{v}_1, \dots, \mathbf{v}_k\}) := \left\{ \sum_{i=1}^k a_i \mathbf{v}_i \mid a_1, \dots, a_k \in \mathbb{R}_{\geq 0} \right\}.$$

2. Their *span* is denoted by

$$\text{Span}(\{\mathbf{v}_1, \dots, \mathbf{v}_k\}) := \left\{ \sum_{i=1}^k a_i \mathbf{v}_i \mid a_1, \dots, a_k \in \mathbb{R} \right\}.$$

Notation II.4. For a matrix $\mathbb{M}$, we denote:

1. the set rows of $\mathbb{M}$ as $\text{rows}(\mathbb{M})$.
2. the set of columns of $\mathbb{M}$ as $\text{cols}(\mathbb{M})$.
3. the i^{th} row as $(\mathbb{M})_i$.
4. the j^{th} column as $(\mathbb{M})^j$.
5. the entry in the i^{th} row and j^{th} column as $(\mathbb{M})_i^j$.

Notation II.5 (Restriction). Let $\mathbb{M}$ denote a $m \times n$ matrix with m rows and n columns. Let $M \subseteq \{1, \dots, m\}$ and $N \subseteq \{1, \dots, n\}$ be a subset of rows and columns, respectively. Then the *restriction* of $\mathbb{M}$ to the rows of M and columns of N is denoted as $\mathbb{M}_M$ and $\mathbb{M}^N$, respectively. The simultaneous restriction of $\mathbb{M}$ to both sets M and N is denoted as

$$\mathbb{M}|_{(M,N)} := (\mathbb{M})_{M,N}^N.$$

Notation II.6 (Support). The *support* of a vector $\mathbf{v}$, denoted $\text{supp}(\mathbf{v})$, is the set of its non-zero coordinates.

B. Chemical reaction networks

Definition II.2. A chemical reaction network (CRN) $\mathcal{G}$ is defined as the triple of a species set, complex set, and reaction set (see [13])

$$\mathcal{G} = (\mathcal{S}, \mathcal{C}, \mathcal{R}),$$

where:

1. the species set $\mathcal{S}$, of size $|\mathcal{S}|$, consists of all the **species** that appear in the network,

$$\mathcal{S} = \{X_1, \dots, X_{|\mathcal{S}|}\}.$$

2. the complex set $\mathcal{C}$ consists of **complexes** of the network. A complex is a formal non-negative integer linear combination of elements of $\mathcal{S}$. The coefficients of the species in a complex form a vector in $\mathbb{Z}_{\geq 0}^{|\mathcal{S}|}$ and denote the *stoichiometry* of the complex. A complex $y \in \mathcal{C}$ will be taken to mean either of the following:

$$\sum_{i=1}^{|\mathcal{S}|} y_i X_i \equiv y \in \mathbb{Z}_{\geq 0}^{|\mathcal{S}|},$$

and be denoted by a column vector.

3. the reaction set $\mathcal{R} \subseteq \mathcal{C} \times \mathcal{C}$ consists of **reactions** of the network. A reaction is an ordered pair of complexes, and has the form

$$\sum_{i=1}^{|\mathcal{S}|} r_i^- X_i \rightarrow \sum_{i=1}^{|\mathcal{S}|} r_i^+ X_i,$$

or, equivalently

$$r^- \rightarrow r^+.$$

Here $r^-, r^+ \in \mathcal{C}$ and are referred to as the *input* and *output* complex, respectively.

Remark 1. Since the information of the set of complexes $\mathcal{C}$ is implicit in the set of reactions $\mathcal{R}$, Deshpande et al. in [9] economically define a CRN simply by the pair $\mathcal{G} = (\mathcal{S}, \mathcal{R})$. Here the complex set is given by taking the union of all the input and output complexes for all the reactions,

$$\mathcal{C} = \bigcup \{r^-, r^+ | r^- \rightarrow r^+ \in \mathcal{R}\}.$$

Remark 2. Isomorphically, a CRN $\mathcal{G} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ can be represented as a directed multi-hypergraph [17, 18] whose hyper-vertices are multisets of vertices according to the dictionary:

	CRN	Hypergraph
$\mathcal{S}$	Species	Vertices
$\mathcal{C}$	Complexes	Hyper-Vertices
$\mathcal{R}$	Reactions	Hyper-Edges

C. Input-output matrix pair and the stoichiometric matrix

Recall that every reaction is of the form

Input complex $\rightarrow$ Output complex.

Notation II.7. We denote a reaction $r \in \mathcal{R}$ as

$$r^- \rightarrow r^+,$$

where r^-, r^+ are column vectors in $\mathbb{Z}_{\geq 0}^{|\mathcal{S}|}$.

For any CRN, we can define $\mathcal{S} \times \mathcal{R}$ matrices $\mathbb{S}^-$ and $\mathbb{S}^+$ by collecting all the input and output complexes, respectively, where

$$(\mathbb{S}^-)^r = r^-, \quad (1)$$

$$(\mathbb{S}^+)^r = r^+. \quad (2)$$

We refer to $\mathbb{S}^-$ and $\mathbb{S}^+$ as the *input* and *output* matrix, respectively. Observe that $\mathbb{S}^-$ and $\mathbb{S}^+$ have only non-negative entries.

Lemma II.1. Each CRN $\mathcal{G}$ has a unique representation as

$$\mathcal{G} = (\mathcal{S}, \mathbb{S}^-, \mathbb{S}^+),$$

where $\mathcal{S}$ denotes the species set and $(\mathbb{S}^-, \mathbb{S}^+)$ represent the input-output matrix pair.

Proof. Using Remark 1, a CRN $\mathcal{G}$ is given by the pair of species and reactions $\mathcal{S}, \mathcal{R}$. By definition, a reaction $r = r^- \rightarrow r^+$ is in the set $\mathcal{R}$ if and only if (iff) the r^{th} columns of $\mathbb{S}^+$ and $\mathbb{S}^-$ are r^+ and r^- , respectively. $\square$

Remark 3. Definition II.2 and the following remarks provide a graph-theoretic (GT) representation of CRNs, and Lemma II.1 provides a linear-algebraic (LA) representation of CRNs.

Definition II.3 (Stoichiometric matrix). The *stoichiometric matrix* $\mathbb{S}$ is a $\mathcal{S} \times \mathcal{R}$ matrix defined as the difference of the output and input matrices,

$$\mathbb{S} = \mathbb{S}^+ - \mathbb{S}^-. \quad (3)$$

Alternatively, for $r \in \mathcal{R}$ and using Eq. 2, each column of the stoichiometric matrix is given by the difference of the output and input vectors of a reaction,

$$(\mathbb{S})^r = r^+ - r^-. \quad (4)$$

Definition II.4 (Stoichiometric subspace). The span of the columns or the image of the stoichiometric matrix is termed as the *stoichiometric subspace*.

$$\text{Stoichiometric subspace} := \text{Image}(\mathbb{S}).$$

D. Hypergraph flows and conservation laws

Definition II.5 (Flow). For a CRN $\mathcal{G} = (\mathcal{S}, \mathcal{R})$, a column vector $\mathbf{v} \in \mathbb{R}_{\geq 0}^{|\mathcal{R}|}$ will be said to define a (hypergraph) flow on the CRN.

A flow on a graph can also be used to create a linear combination of the reactions, also called a composite reaction in [6]. While, technically, an abuse of notation, let us use r to index the reaction set $\mathcal{R}$ and also refer to the r^{th} reaction. Then, a flow $\mathbf{v} = [v_1, \dots, v_r, \dots, v_{|\mathcal{R}|}]^T$ on CRN $\mathcal{G}$ would result in the composite reaction

$$\sum_{r \in \mathcal{R}} v_r r^- \rightarrow \sum_{r \in \mathcal{R}} v_r r^+.$$

Let vector $\mathbf{n} = [n_1, \dots, n_{|\mathcal{S}|}]^T \in \mathbb{Z}_{\geq 0}^{|\mathcal{S}|}$ denote the population of each species in the system, and $\mathbf{v} \in \mathbb{Z}_{\geq 0}^{\mathcal{R}}$ denote a flow on $\mathcal{G}$.¹ The amount of species consumed and produced due to this flow are given by $\mathbb{S}^- \cdot \mathbf{v}$ and $\mathbb{S}^+ \cdot \mathbf{v}$, respectively. Thus, the change of species population $\Delta \mathbf{n}$ resulting from the flow $\mathbf{v}$ on the graph $\mathcal{G}$ is given by

$$\Delta \mathbf{n} = \mathbb{S} \cdot \mathbf{v}.$$

Note that, from Definition II.4, the change in species population under an arbitrary flow on $\mathcal{G}$ lies in the stoichiometric subspace.

The right null space of $\mathbb{S}$ correspond to *null flows* which leave the population of the species unchanged ($\Delta \mathbf{n} = 0$). An accounting of the right null space of $\mathbb{S}$ provides a topological classification for CRNs, which we review in Appendix A.

Definition II.6 (Null flows). A column vector $\mathbf{w} \in \mathbb{R}^{|\mathcal{R}|}$ is said to be a null flow on the CRN $\mathcal{G}$ if

$$\mathbb{S} \cdot \mathbf{w} = \mathbf{0}.$$

The left null space of $\mathbb{S}$, i.e. vectors $\mathbf{x} \in \mathbb{R}^{|\mathcal{S}|}$ such that $\mathbf{x} \cdot \mathbb{S} = \mathbf{0}$, correspond to *conservation laws*

$$\mathbf{x} \cdot \Delta \mathbf{n} = 0.$$

Equivalently, the inner product of the conservation law with the population $\mathbf{x} \cdot \mathbf{n}$ yields a *conserved quantity* that is an invariant for the network under an arbitrary flow.

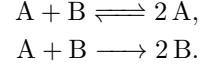
Definition II.7 (Conservation law). For a CRN $\mathcal{G} = (\mathcal{S}, \mathcal{R})$, a column vector $\mathbf{x}$ of size $|\mathcal{S}|$ will be said to be a conservation law if

$$\mathbf{x}^T \cdot \mathbb{S} = \mathbf{0}.$$

¹ We will use a notation of counts for definiteness and simplicity. A parallel construction goes through with concentrations and flow vectors $\in \mathbb{R}_{\geq 0}^{|\mathcal{S}|}$ and can be found in [19].

The vector $\mathbf{x}$ will be called a *positive conservation law* or *mass-like conservation law* if all its entries are nonnegative. Moreover, if the entries are nonnegative integers, it will be called a *positive integer valued conservation law*.

Example II.1. Consider the CRN $\mathcal{G} = (\mathcal{S}, \mathcal{R})$ given as



For a graphical representation of this example, see Fig. 2. Then the species set $\mathcal{S} = \{A, B\}$ and the reaction set $\mathcal{R} = \{A + B \rightarrow 2A, 2A \rightarrow A + B, A + B \rightarrow 2B\}$. As can be read from $\mathcal{R}$, the complex set $\mathcal{C} = \{A + B, 2A, 2B\}$, and their corresponding vectors are

$$y_{A+B} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, y_{2A} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, y_{2B} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

The input and output matrices, $\mathbb{S}^-$ and $\mathbb{S}^+$, respectively, for $\mathcal{G}$ are

$$\mathbb{S}^- = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix},$$

and

$$\mathbb{S}^+ = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}.$$

The stoichiometric matrix $\mathbb{S}$ for $\mathcal{G}$ is

$$\mathbb{S} = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \end{bmatrix},$$

with left null space (conservation laws) spanned by $\mathbf{x} = [1, 1]$ and right null space (null flows) spanned by the basis $\{[1, 1, 0]^T, [1, 0, 1]^T\}$. Notice that the stoichiometric subspace is one-dimensional and the Euclidean inner-product of the spanning vectors with the conservation law is zero.

E. Formal, exclusive and stoichiometric autocatalysis

Consider a CRN $\mathcal{G} = (\mathcal{S}, \mathcal{R})$ with its stoichiometric matrix $\mathbb{S}$.

Definition II.8 (Subnetwork). A CRN $\mathcal{G}' = (\mathcal{S}', \mathcal{R}')$ is a subnetwork of $\mathcal{G}$ if:

1. The reaction set of $\mathcal{G}'$ is a subset of $\mathcal{G}$,

$$\mathcal{R}' \subseteq \mathcal{R}.$$

2. Every species that appears in the subset $\mathcal{R}'$ in $\mathcal{G}$ appears in $\mathcal{S}'$. Equivalently,

$$X_i \in \mathcal{S}' \text{ iff } r^- \rightarrow r^+ \in \mathcal{R}' \text{ such that } r_i^- > 0 \text{ or } r_i^+ > 0.$$

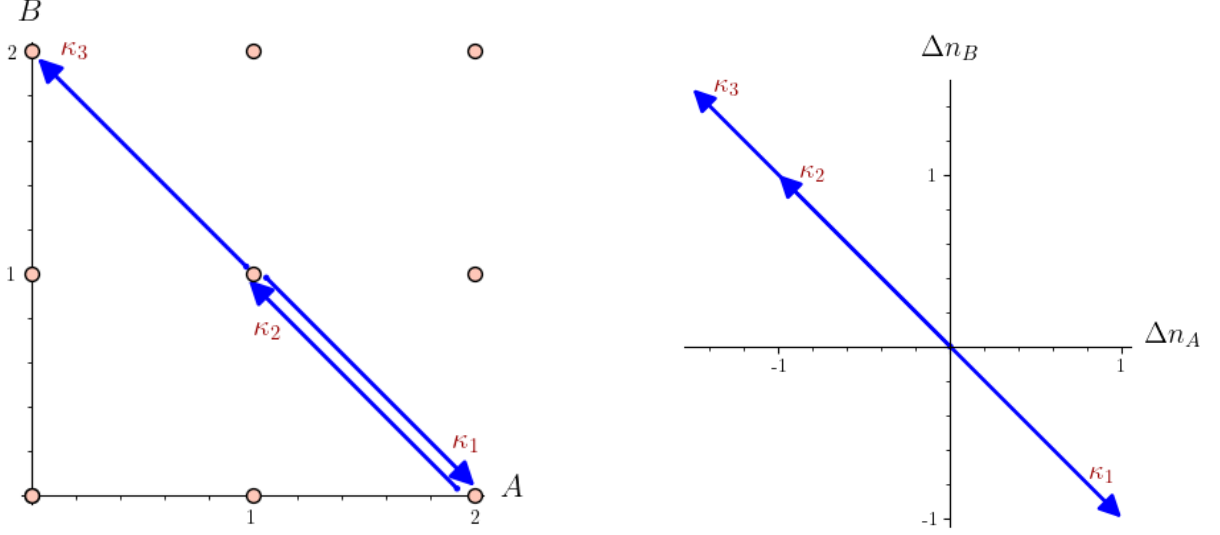


Figure 2: In the left panel, the Euclidean embedded graph (E-graph, see [20]) of the CRN in Example II.1 is shown. The E-graph is obtained by representing the complexes as lattice points in a Euclidean lattice and representing the reactions as edges between them. The flows on the reactions are labelled by κ and the change in concentration they induce, $\Delta n = S\kappa$, is schematically represented in the right panel.

Remark 4. The stoichiometric matrix of the subnetwork is obtained from that of the network by keeping only the columns and rows corresponding to the reactions and species participating in the subnetwork, respectively.

Definition II.9 (Motif). A CRN $\mathcal{G}' = (\mathcal{S}', \mathcal{R}')$ is a motif (or a subhypergraph) of $\mathcal{G}$ if (see [11]):

1. The reaction set of $\mathcal{G}'$ is a subset of the reaction set of $\mathcal{G}$,

$$\mathcal{R}' \subseteq \mathcal{R}.$$

2. The species set of $\mathcal{G}'$ is a subset of the species set of $\mathcal{G}$,

$$\mathcal{S}' \subseteq \mathcal{S}.$$

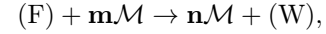
Notation II.8. The stoichiometric matrix of a motif is obtained from that of the network by keeping only the rows and columns corresponding the species and reactions in the motif, respectively. Henceforth, we denote an arbitrary submatrix of the stoichiometric matrix corresponding to the motif $(\mathcal{S}', \mathcal{R}')$ with an overbar,

$$\bar{S} := S|_{(\mathcal{S}', \mathcal{R}')} = (S)_{\mathcal{S}'}^{\mathcal{R}'}.$$

1. *Formal autocatalysis*

Definition II.10. A subnetwork $\mathcal{G}' = (\mathcal{S}', \mathcal{R}')$ of CRN $\mathcal{G} = (\mathcal{S}, \mathcal{R})$ is defined to be *formally autocatalytic* in the subset $\mathcal{M}$ of $\mathcal{S}'$ if (see [6, 21]):

1. There exists a positive real flow ($\gg 0$) on $\mathcal{G}'$ such that the resulting composite reaction is of the form



where $\mathbf{0} \ll \mathbf{m} \ll \mathbf{n}$. Here $\mathbf{m}$ and $\mathbf{n}$ are the stoichiometries of the set $\mathcal{M}$ in the input and output of the composite reaction, respectively, and $\mathbf{o}\mathcal{M} = \sum_i \mathbf{o}_i \mathcal{M}_i$.

Remark 5. This definition can be generalized to semi-positive flows, but we will not need that generalization for this work. (Also, see Remark 11.)

Lemma II.2. A subnetwork $\mathcal{G}' = (\mathcal{S}', \mathcal{R}')$ of CRN $\mathcal{G} = (\mathcal{S}, S^-, S^+)$ is formally autocatalytic in the subset $\mathcal{M}$ iff:

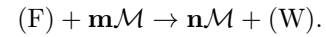
1. There exists a flow $\mathbf{v} \gg \mathbf{0}$ on $\mathcal{G}'$ such that

$$\bar{S} \cdot \mathbf{v} = (S)_{\mathcal{M}}^{\mathcal{R}'} \cdot \mathbf{v} \gg \mathbf{0}.$$

2. All species in $\mathcal{M}$ are consumed in at least some reaction in $\mathcal{R}'$, or

$$w > \mathbf{0} \text{ for each } w \in \text{rows}((S^-)_{\mathcal{M}}^{\mathcal{R}'}).$$

Proof. Let the flow $\mathbf{v} \gg \mathbf{0}$ on the subnetwork $\mathcal{G}'$ result in a composite reaction of the form



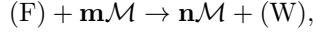
Condition 1 then yields that

$$\mathbf{n} - \mathbf{m} \gg \mathbf{0},$$

and condition 2 implies that

$$\mathbf{m} \gg \mathbf{0}.$$

Thus, the conditions stated in the lemma imply that the subnetwork is formally autocatalytic. Conversely, suppose there is a flow $\mathbf{v}$ that leads to the composite reaction of the form



such that $\mathbf{0} \ll \mathbf{m} \ll \mathbf{n}$. Then,

$$\mathbf{m} \gg \mathbf{0} \implies \text{rows}((\mathbb{S}^-)_{\mathcal{M}}) > \mathbf{0},$$

and

$$\mathbf{n} - \mathbf{m} \gg \mathbf{0} \implies \mathbb{S} \cdot \mathbf{v}|_{\mathcal{M}} \gg \mathbf{0}.$$

□

Remark 6. A matrix $\bar{\mathbb{S}} = \mathbb{S}|_{\mathcal{M}}$ for which there exists a $\mathbf{v} \gg \mathbf{0}$ such that

$$\bar{\mathbb{S}} \cdot \mathbf{v} \gg \mathbf{0}$$

is also referred to as a *productive* matrix in [11, Supplementary Information] and is known as a *semi-positive* matrix [22, Theorem 3.1.13]. Notice that, productivity implies

$$\text{rows}((\mathbb{S}^+)_{\mathcal{M}}) > \mathbf{0}.$$

Remark 7. The productivity condition can be restated as the condition that the interior of the cone induced by the columns of the stoichiometric matrix intersects the positive orthant in the autocatalytic species,

$$\text{interior}(\text{cone}(\text{cols}((\mathbb{S})))) \cap \mathbb{R}_{\geq 0}^{\mathcal{M}} \neq \emptyset.$$

Remark 8. The pair $(\mathcal{M}, \mathcal{R}')$ defines an autocatalytic motif.

Example II.2. Consider the reaction network $\mathcal{G}_f = (\{A, B\}, \{A \rightarrow B, A + B \rightarrow B\})$. Under the flow $\mathbf{v} = [1, 1]^T$, the resulting composite reaction is



Correspondingly, the input and stoichiometric matrix for the CRN are

$$\mathbb{S}^- = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

and

$$\mathbb{S} = \begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix},$$

respectively. It is easy to check that both requirements of Lemma II.2 hold true in the species set $\{B\}$. Thus the reaction network is formally autocatalytic for the set $\mathcal{M} = \{B\}$.

2. Exclusive autocatalysis

Definition II.11 (GT). A formally autocatalytic subnetwork $\mathcal{G}' = (\mathcal{S}', \mathcal{R}')$ of CRN $\mathcal{G} = (\mathcal{S}, \mathcal{R})$ is defined to be *exclusively autocatalytic* in the subset $\mathcal{M}$ (see [6, 9, 23, 24]) if:

- Every reaction in $\mathcal{R}'$ consumes at least one species from the set $\mathcal{M}$. This ensures that the flow is inadmissible, or there is no flow, if the population of any species in the set $\mathcal{M}$ is zero.

Lemma II.3. A formally autocatalytic subnetwork $\mathcal{G}' = (\mathcal{S}', \mathcal{R}')$ of CRN $\mathcal{G} = (\mathcal{S}, \mathbb{S}^-, \mathbb{S}^+)$ is *exclusively autocatalytic* in the subset $\mathcal{M}$ if and only if:

- The restriction of the input complexes to the set $\mathcal{M}$ is greater than zero, or

$$r^- > \mathbf{0} \text{ for each } r^- \in \text{cols}((\mathbb{S}^-)^{\mathcal{R}'}_{\mathcal{M}}).$$

Proof. It is clear that if the input complex $r^- > \mathbf{0}$ for each $r^- \in \text{cols}((\mathbb{S}^-)_{\mathcal{M}})$, then each reaction consumes at least one species from the set $\mathcal{M}$, and vice versa. □

Remark 9. If, for a reaction $r^- \rightarrow r^+$ in the exclusively autocatalytic subnetwork, the output complex has no species in the set $\mathcal{M}$, then such a reaction can be removed from the subnetwork to yield a smaller exclusively autocatalytic subnetwork. Thus, ‘smaller’ (not necessarily minimal) autocatalytic subnetworks can be obtained by further imposing the condition:

$$r^+ > \mathbf{0} \text{ for each } r^+ \in \text{cols}((\mathbb{S}^+)_{\mathcal{M}}^{\mathcal{R}'}).$$

Remark 10. For a reaction network, a species set $\mathcal{M}$ that satisfies

$$\text{cols}((\mathbb{S}^+)_{\mathcal{M}}) > \mathbf{0}$$

and

$$\text{cols}((\mathbb{S}^-)_{\mathcal{M}}) > \mathbf{0}$$

is termed as a *siphon* by Deshpande et al. in [9]. Moreover, if the network is also productive, the set $\mathcal{M}$ is termed as a self-replicating siphon.

Example II.3. One can verify that the CRN $\mathcal{G}$ in Example II.1 is:

1. Exclusively autocatalytic in the set $\mathcal{M} = \{A\}$.
2. Not exclusively autocatalytic in the sets $\mathcal{M} = \{B\}$ or $\{A, B\}$.

Similarly, one can verify that the CRN $\mathcal{G}_f$ in Example II.2 is not exclusively autocatalytic since the flow proceeds even if the population of B is 0.

3. Stoichiometric autocatalysis

Definition II.12. A subnetwork $\mathcal{G}' = (\mathcal{S}', \mathcal{R}')$ of CRN $\mathcal{G} = (\mathcal{S}, \mathcal{R})$ is *stoichiometrically autocatalytic* in the subset $\mathcal{M}$ if:

1. It is exclusively autocatalytic.
2. For every reaction $r^- \rightarrow r^+ \in \mathcal{R}'$:
 - (a) each of r^- and r^+ contain at least one species from the set $\mathcal{M}$.
 - (b) no species is in both r^- and r^+ ,

$$\text{supp}(r^-) \cap \text{supp}(r^+) = \emptyset.$$

Lemma II.4. An exclusively autocatalytic subnetwork $\mathcal{G}' = (\mathcal{S}', \mathcal{R}')$ of CRN $\mathcal{G} = (\mathcal{S}, \mathbb{S}^-, \mathbb{S}^+)$ is *stoichiometrically autocatalytic in the subset $\mathcal{M}$ only if*:

1. Each column in $\bar{\mathbb{S}} = (\mathbb{S}^+ - \mathbb{S}^-)_{\mathcal{M}}$ has at least one positive and one negative entry.
2. Each row in $\bar{\mathbb{S}}$ has at least one positive and one negative entry.

Proof. The conditions in Definition II.12 implies that every column in the stoichiometric matrix $\mathbb{S}^+ - \mathbb{S}^-$ restricted to the autocatalytic set contains a positive and a negative entry. From condition 2 in Lemma II.2 and Remark 6, we know that every row in the input and output matrix pair restricted to the autocatalytic set is semi-positive, respectively. This, along with the condition that every reaction has distinct input and output species, yields that every row in the stoichiometric matrix restricted to the autocatalytic set must contain a positive and a negative entry. $\square$

Remark 11. The preceding lemma alludes to the equivalence of our notion of stoichiometric autocatalysis and the autocatalysis of Blokhuis et al. in [11]. They term the first condition as *autonomy* of the submatrix $\bar{\mathbb{S}}$. The second condition implies that each autocatalytic species is consumed (and produced). However, there is a caveat. The productivity condition for Blokhuis et al. is defined over an arbitrary flow vector which can take negative values. In their treatment, they only consider reversible reaction networks where for a reaction with a negative flow, an opposite reaction with a positive flow exists. Since, we do not make such assumptions on our reaction network, we restrict our flows to be strictly in the positive orthant.

The condition $\text{cols}((\mathbb{S}^-)_{\mathcal{M}}) > 0$, ensures that the production of any species *consumes* at least one other species of the motif, disallowing unconditional growth. Notice that this condition is satisfied by exclusive and stoichiometrically autocatalytic subnetworks. However, the condition that the supports of the input and output complexes are disjoint makes the definition of stoichiometric

autocatalysis more restrictive than exclusive autocatalysis. A useful feature of stoichiometric autocatalysis is that it can be inferred from the stoichiometric matrix without referring to the underlying CRN, due to which their criterion is termed stoichiometric autocatalysis [25].

Following the terminology in [9], if there exists a non-negative flow such that $\bar{\mathbb{S}} \cdot \mathbf{v} \ll \mathbf{0}$, we would term the motif $(\mathcal{M}, \mathcal{R}')$ a *drainable motif*, where the change in concentration of all the species in the subset $\mathcal{M}$ is strictly negative. While we focus our attention on autocatalytic motifs, our results are easily extended to their drainable counterparts. In [9], the authors prove that a network is persistent if it has no drainable motifs and also remark on the importance of autocatalytic and drainable motifs for understanding the dynamics of a network in ecological terms.

III. MINIMAL AUTOCATALYTIC SUBNETWORKS

In the previous section, we gave the criteria for ascertaining when a subnetwork is exclusively autocatalytic. However, one autocatalytic subnetwork can have several subnetworks which are also autocatalytic. In Sec. III A, we define and prove some properties of minimal autocatalytic subnetworks (MASs). In Sec. III B, we demonstrate that these subnetworks organize within equivalence classes. Finally, in Sec. III C we investigate how the subnetworks within an equivalence class can be organized by analyzing what we define as their *flow-productive*, *species-productive* and *partition-productive cones*.

A. Properties

To define a MAS, we first define an **autocatalytic core** [11]. Recall from Definition II.9 that a motif is an arbitrary subhypergraph which is obtained by selecting a subset of the reactions and species in the original CRN.

Definition III.1 (Autocatalytic core). An *autocatalytic core* is an exclusively autocatalytic motif that does not contain a smaller exclusively autocatalytic motif.

Remark 12. Note that our definition is a generalization of an *autocatalytic core* as defined by Blokhuis et al. as ours is defined for *exclusively*, as opposed to *stoichiometrically*, autocatalytic motifs.

Notation III.1. We denote the motif of the autocatalytic core $A = (A_C, \mathcal{R}_A)$, its input-output matrix pair as $(\bar{\mathbb{S}}^+, \bar{\mathbb{S}}^-)$, and term A_C as the *core set*.

Theorem III.1. A motif $(A_C, \bar{\mathbb{S}}^+, \bar{\mathbb{S}}^-)$ is an autocatalytic core only if all the following conditions are satisfied:

1. The stoichiometric matrix $\bar{\mathbb{S}} = \bar{\mathbb{S}}^+ - \bar{\mathbb{S}}^-$ is productive, that is

$$\exists \mathbf{v} \gg \mathbf{0} \text{ such that } \bar{\mathbb{S}} \cdot \mathbf{v} \gg \mathbf{0}.$$

2. Every species in the core set is consumed at least once,

$$\text{rows}(\overline{\mathbb{S}}^-) > \mathbf{0}.$$

3. Every species in the core set is produced at least once,

$$\text{rows}(\overline{\mathbb{S}}^+) > \mathbf{0}.$$

4. Every reaction in the motif consumes some species in the core set,

$$\text{cols}(\overline{\mathbb{S}}^-) > \mathbf{0}.$$

5. Every reaction in the motif produces some species in the core set,

$$\text{cols}(\overline{\mathbb{S}}^+) > \mathbf{0}.$$

Proof. First, notice that every species in the autocatalytic core must be in the autocatalytic set, otherwise it could be removed to yield a smaller autocatalytic motif. Using Lemma II.3, properties 1, 2, and 4 follow. As remarked in 6, property 3 follows from property 1. Finally, as remarked in 9, if a reaction does not produce species in the autocatalytic set, it could be removed without affecting productivity. $\square$

Furthermore, we can easily extend the proof of Proposition 2 in the Supplementary Information for [11] by Blokhuis et al. to show that (exclusively) autocatalytic cores are always square and invertible.

Lemma III.2. *Every species in the core set is the only reactant of at least one reaction in the autocatalytic core.*

Proof. Let there be a species in the core set that is never the only reactant of any reaction and always occurs as a co-reactant. Since this species is in an autocatalytic core, it must be produced by some reactions. This species occurs either as:

1. an only product of reactions.
2. a co-product of reactions.

If the species under consideration occurs as an only product of any reaction, then this reaction can be removed to yield a smaller motif which is autocatalytic in the remaining species. If the species only occurs as a co-product of a reaction alongside other species, then the species under consideration can be removed to yield a smaller autocatalytic motif. In either case, minimality of the autocatalytic core is violated, leading to a contradiction. $\square$

Recall from Remark 6 that the stoichiometric matrix of an exclusively autocatalytic motif is semi-positive. A semi-positive matrix is *minimal semi-positive* [22, Section 3.5] if no column can be removed resulting in a semi-positive matrix.

Lemma III.3. *The stoichiometric matrix of an autocatalytic core is a minimal semi-positive matrix.*

Proof. Removing a column of the stoichiometric matrix corresponds to removing a reaction from the motif.

Suppose for the sake of contradiction that removing a column of the stoichiometric matrix of an autocatalytic core results in a motif with semi-positive stoichiometric matrix.

Then removing the species not consumed by any of the remaining reactions results in a smaller motif that also has a semi-positive stoichiometric matrix, and in which every species is consumed by some reaction. Moreover, any reaction consuming a species in the original motif still consumes a species in the smaller motif, so the resulting motif is exclusively autocatalytic, contradicting minimality of the core. $\square$

Theorem III.4. *The stoichiometric matrix of an autocatalytic core is a square and invertible matrix. In particular, the number of reactions in autocatalytic core equals the number of species in the core set.*

Proof. By Lemma III.3, the stoichiometric matrix of an autocatalytic core is a minimal semi-positive matrix, and so has linearly independent columns whose number is bounded by the number of rows [22, Theorem 3.5.6]. Lemma III.2 shows the number of species, i.e. rows, is bounded by the number of reactions, i.e. columns. Thus the stoichiometric matrix is square with linearly independent columns, and hence also invertible. $\square$

Remark 13. The autocatalytic cores of [11] are autocatalytic cores in our sense that are stoichiometrically autocatalytic and satisfy the additional property that no smaller motif becomes autocatalytic after reversing any of its reactions. Blokhuis et al. provide a graphical characterization of the stoichiometrically autocatalytic cores and show that they can only be of five types. In Fig. 3 we have listed all the minimal autocatalytic subnetworks for the complete 1-constituent CCRN (up to a maximum cluster size of 4) and have color coded them by their types.

In a CRN, an autocatalytic core will generally occur embedded in a set of reactions. From Definition II.8, recall that a subnetwork is obtained from a CRN by selecting a subset of reactions and retaining all the species that participate in the subset. We define **minimal autocatalytic subnetworks** as the following.

Definition III.2 (MAS). A *minimal autocatalytic subnetwork* (MAS) is defined to be the subnetwork with the least number of reactions containing a particular autocatalytic core.

Remark 14. For a particular autocatalytic core, a MAS loses its autocatalytic property over all the core species if any reaction is removed. Moreover, if any reaction is added, the subnetwork is not minimal. Thus, a MAS contains exactly the number of reactions as its particular core.

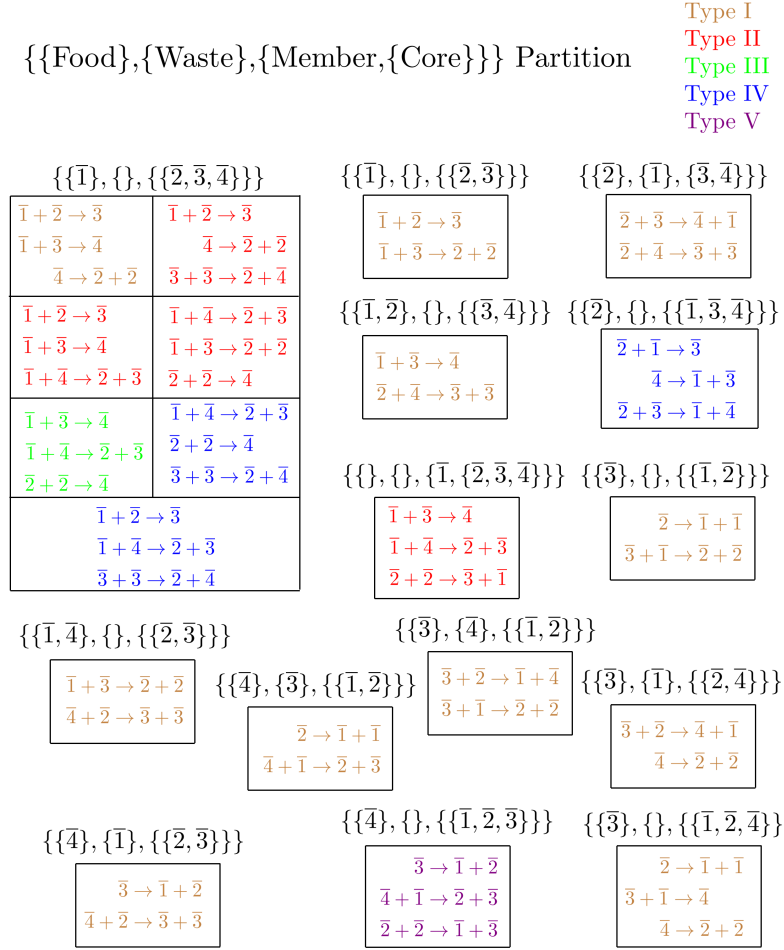


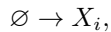
Figure 3: All minimal autocatalytic subnetworks (MASs) for the complete 1-constituent CCRN with $L = 4$ are shown. The core types are also color coded using the typology of Blokhuis et al. in [11]. In a CCRN, each species is denoted by a vector of positive integers with an overbar, and each reaction satisfies certain conservation laws. For a definition of a CCRN, see Sec. V. The list of reactions can be found in Appendix E Table III.

Remark 15. A MAS can also contain two distinct but overlapping autocatalytic cores. A simple example of such a network is:

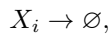
$$\mathcal{G} = \{A + B \rightarrow C, C \rightarrow 2A + 2B\}.$$

Observe that, while $\mathcal{G}$ is a MAS, it contains two autocatalytic cores consisting of autocatalytic species sets $\{A, C\}$ and $\{B, C\}$.

Remark 16. Observe that, using Theorem III.1, an *inflow reaction* of the type



or an *outflow reaction* of the type



can never be in the reaction set of a MAS. We return to this point in Sec. IV A.

Example III.1. Consider the CRN $\mathcal{G}$ given by the reaction set

$$\mathcal{R} = \{A \rightarrow A + B, B \rightarrow A\}.$$

The input-output matrix pair and the stoichiometric matrix for this CRN are given by:

$$\mathbb{S}^- = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$\mathbb{S}^+ = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix},$$

$$\mathbb{S} = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}.$$

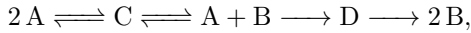
It can be verified that this CRN is exclusively autocatalytic but *not* stoichiometrically autocatalytic. Note,

moreover, consistent with Theorem III.4, the stoichiometric matrix is square and invertible.

Example III.2. Recall Example II.1 with the stoichiometric matrix

$$\mathbb{S} = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 1 \end{bmatrix}.$$

It can be verified that this CRN is exclusively autocatalytic but not stoichiometrically autocatalytic. To convert this to a stoichiometrically autocatalytic network, we can modify the network by replacing each reaction whose reactant and product complex shares a species with two new reactions with a fictitious distinct intermediate species. This then yields a new network $\mathcal{G}^*$ given by



where C, D are the newly added fictitious species. The stoichiometric matrix for $\mathcal{G}^*$ is given by

$$\mathbb{S}^* = \begin{bmatrix} 2 & -2 & -1 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 & -1 & 2 \\ -1 & 1 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}.$$

Notice that the restriction of the stoichiometric matrix to the species set $\mathcal{S}' = \{A, C\}$ and the reactions set $\mathcal{R}' = \{A + B \rightarrow C, C \rightarrow 2A\}$, denoted by $(\mathbb{S}^*)_{\mathcal{S}'}^{\mathcal{R}'}$, indeed satisfies the properties for an autocatalytic core, and thus $\mathcal{G}^*$ is indeed autocatalytic with core species $\{A, C\}$. A similar construction shows that the network is also autocatalytic in the set $\{B, D\}$. (In the typology of Blokhuis et al. both of these cores are of type I.)

B. Food-waste-member-core partition

Let $\mathcal{G} = (\mathcal{S}, \mathcal{R})$ be a CRN and $A = (\mathcal{S}_A, \mathcal{R}_A)$ be a MAS autocatalytic in the core species A_C . From Theorem III.4, we know that the number of core species is the same as the size of $\mathcal{R}_A$. In general, however, there will also be disjoint subsets of $\mathcal{S}_A$ of species that only occur as co-reactants, co-products, and both co-reactants and co-products in $\mathcal{R}_A$. As explained below, given a MAS and an associated autocatalytic core, we can partition the species set into *food-waste-member-core* (FWMC) sets.

Definition III.3 (FWMC). Let the stoichiometric matrix of MAS $A = (\mathcal{S}_A, \mathcal{R}_A)$ with a particular core be denoted by $\bar{\mathbb{S}}$. Denoting the i^{th} row of a matrix $\mathbb{M}$ as $(\mathbb{M})_i$, we define:

1. the species set corresponding to all rows of $\bar{\mathbb{S}}$ with nonpositive coefficients as the *food set* and denote it by A_F ,

$$X_i \in A_F \text{ iff } (\bar{\mathbb{S}})_i < \mathbf{0}.$$

2. the species set corresponding to all rows of $\bar{\mathbb{S}}$ with nonnegative coefficients as the *waste set* and denote it by A_W ,

$$X_i \in A_W \text{ iff } (\bar{\mathbb{S}})_i > \mathbf{0}.$$

3. the species set corresponding to all rows of $\bar{\mathbb{S}}$ with both positive and negative coefficients as the *member set* and denote it by A_M .
4. the species set corresponding to all rows of $\bar{\mathbb{S}}$ as the *core member set* and denote it by A_C . A_C is a subset of A_M .

Remark 17. A MAS with distinct, but overlapping, autocatalytic cores (for example, see Remark 15), can have a non-unique FWMC partition.

Notation III.2. We will refer to the submatrices of the stoichiometric matrix generated by restricting to the food, waste, member, and core species, as $\bar{\mathbb{S}}_F$, $\bar{\mathbb{S}}_W$, $\bar{\mathbb{S}}_M$, and $\bar{\mathbb{S}}_C$, respectively.

Notation III.3. We denote the exclusion of a subset K from the species set by $A_{/K}$.

For example, all species in the species set of an autocatalytic subnetwork A except the core member species will be denoted by $A_{/C}$. Note that the core member set, or simply *core set*, is a subset of the member set, or $A_C \subseteq A_M$. We refer to the species in the member set that are not in the core set as *non-core member species* and denote their set as $A_{M/C}$. For an example of the partitioning process on a reaction network where each set is non-empty, see Fig. 4.

We refer to the above partition of the species set for a particular autocatalytic core in a MAS as the *food-waste-member-core* (FWMC) partition of the MAS. Notice that equality of the FWMC partitions of multiple autocatalytic cores can serve as the basis for an equivalence relation.

Definition III.4. Let a and b be two autocatalytic cores embedded in MASs A and B . The autocatalytic cores a and b are equivalent if and only if the FWMC partitions of A and B under their associated autocatalytic cores is identical.

$$a \equiv b \iff \text{FWMC}(A) = \text{FWMC}(B).$$

Remark 18. In the rest of the work, unless specified otherwise, we will assume that a MAS has a unique autocatalytic core. Under this assumption, each MAS can be uniquely assigned its FWMC partition.

Our *quadpartite* partition of the species set can be coarse-grained to a bipartite *resource-member* partition where the *resource* set is the union of food and waste sets (and member set is the same as above) (see Fig. 4). Our partitioning of the species set by examining the entries of the stoichiometric matrix is cognate with the

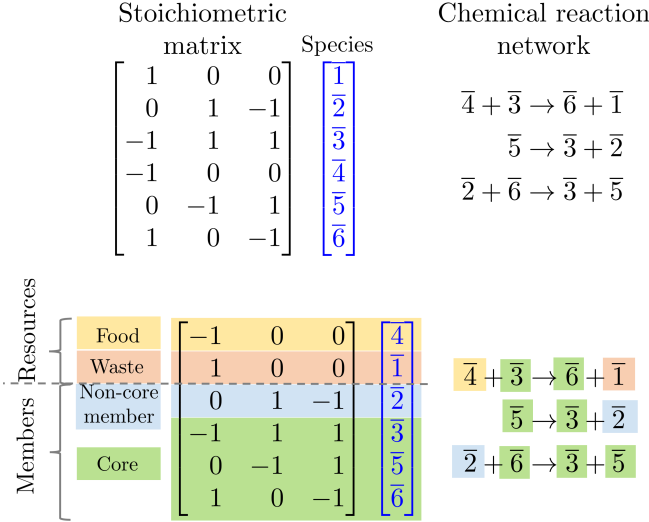


Figure 4: Food-waste-member-core and resource-member partition of a minimal autocatalytic subnetwork in the 1-constituent Cluster CRN (CCRN) with $L = 6$. Using the typology of Blokhuis et al. this subnetwork has a type two core.

partitioning done by Avanzini et al. in [26], where they partition the set of species into a *resource-member* partition. Also, it is a refinement of, the *food-waste-member* partitioning by Peng et al. in [7], where it is shown that different interactions between cycles can be defined based on which types of species are in common. For example, competition applies when two MASs share a food species or a waste species, mutualism, when the food of one is the waste of another, and predation/parasitism, when a member of one is the food of another (see Fig. 1).

Example III.3. In the network $\mathcal{G}^*$ from Example III.2, the MASs are:

1. $A_1 = (\{A, B, C\}, \{A + B \rightarrow C, C \rightarrow 2A\})$ with FWMC partition $\{\{B\}, \{\}, \{\{A, C\}\}\}$.
2. $A_2 = (\{A, B, D\}, \{A + B \rightarrow D, D \rightarrow 2B\})$ with FWMC partition $\{\{A\}, \{\}, \{\{B, D\}\}\}$.

Notice that both these subnetworks have empty waste and non-core member species sets.

C. Flow, species and partition productive cones

Recall from Notation III.2 and Theorem III.4 that $\bar{\mathbb{S}}_C$ is invertible. The chemical interpretation of the inverse is that the k^{th} column is a flow vector that increases the k^{th} species by exactly one unit, making it an elementary mode of production.

Theorem III.5. *The columns of $\bar{\mathbb{S}}_C^{-1}$ lie in the semi-positive orthant.*

Proof. The stoichiometric matrix is a square minimal semi-positive matrix by Lemma III.3 and Theorem III.4, so its inverse has non-negative entries by [22, Corollary 3.5.8], and the columns are non-zero non-negative so semi-positive. $\square$

Flow-productive cone:

Definition III.5. The *flow-productive cone*, denoted by $\mathcal{F}$, is defined to be the cone generated by the elementary modes of production of an autocatalytic core in a MAS,

$$\mathcal{F}(A) = \text{cone}(\text{cols}(\bar{\mathbb{S}}_C^{-1})). \quad (5)$$

Remark 19. Using Theorem III.5, the inverse of $\bar{\mathbb{S}}_C$ never contains negative entries. Thus $\mathcal{F}(A)$ lies in the non-negative orthant.

By definition, an element in the interior of the flow-productive cone, the *flow productive region*, of A corresponds to a flow vector for which A is productive, i.e. the core member species of A are strictly produced.

Species-productive cone:

Definition III.6. The *species-productive cone*, denoted by $\mathcal{P}$, is defined to be the image of the flow-productive cone of an autocatalytic core in a MAS under its stoichiometric matrix $\bar{\mathbb{S}}$,

$$\begin{aligned} \mathcal{P}(A) &= \bar{\mathbb{S}} \cdot \mathcal{F}(A) \\ &= \text{cone}(\text{cols}(\bar{\mathbb{S}} \cdot \bar{\mathbb{S}}_C^{-1})). \end{aligned} \quad (6)$$

Therefore, each vector in the interior of the species-productive cone for a MAS, which we refer to as the *species-productive region*, describes a change in concentration of the participating species for which the subnetwork is productive in the core member species.

Remark 20. The *species-consumptive cone* for the *drainable subnetwork* A^{op} , which is obtained by reversing the edges of the autocatalytic subnetwork A , is the opposite of the species-productive cone:

$$\mathcal{P}(A^{\text{op}}) = -\mathcal{P}(A).$$

Partition-productive cone: Consider a MAS with species set $\mathcal{S}$. Let us denote the FWMC partition of $\mathcal{S}$ as $T = \{T_F, T_W, T_{M/C}, T_C\}$. Let e_i be a column vector of size $\mathcal{S}$ with 1 at the i^{th} position and 0 elsewhere. We define the set of basis vectors $\mathcal{B}_T$ for the partition T to be

$$\mathcal{B}_T = \mathcal{B}_T^F \cup \mathcal{B}_T^W \cup \mathcal{B}_T^{M/C} \cup \mathcal{B}_T^C \quad (7)$$

where

$$\begin{aligned} \mathcal{B}_T^F &= \{-e_f | f \in T_F\}, \\ \mathcal{B}_T^W &= \{e_w | w \in T_W\}, \\ \mathcal{B}_T^{M/C} &= \{-e_m \cup e_m | m \in T_{M/C}\}, \\ \mathcal{B}_T^C &= \{e_c | c \in T_C\}. \end{aligned}$$

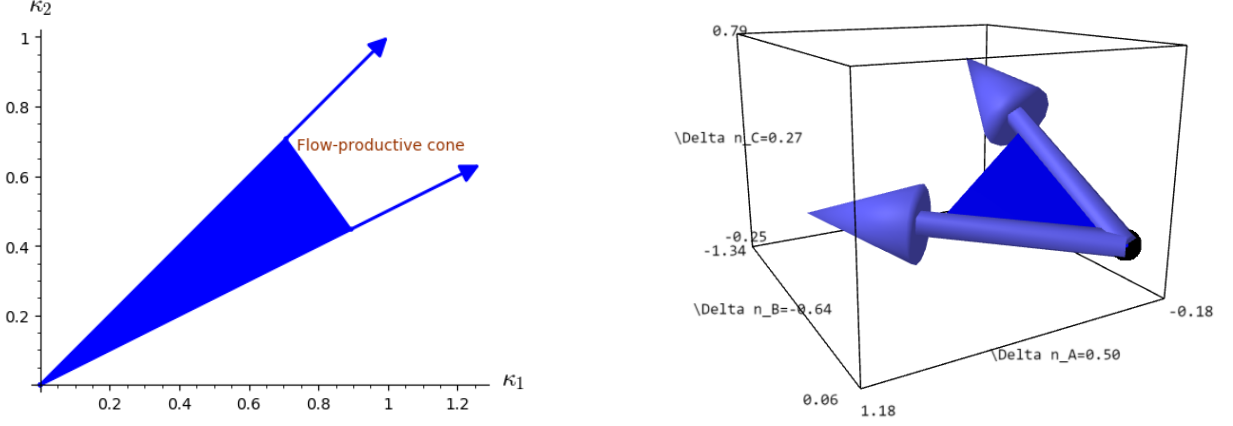


Figure 5: The flow and species productive cones for the Example III.4 are shown in the left and right panel, respectively. The dimensions of the embedding space of the flow and species productive cones are the number of reactions and species in the minimal autocatalytic subnetwork, respectively. Notice that the food species B is strictly consumed (nonpositive) and the core species A, C are strictly produced (nonnegative) in the species-productive region. For this example, the partition-productive cone is identical to the species-productive cone.

Thus the basis $\mathcal{B}_T$ for a partition T spans the negative half-space of each species in the food set, the positive half-space of each species in the waste and core set, and both negative and positive half-spaces of each species in the non-core member set. Let the reaction network have conservation laws $\{\mathbf{x}_i\}$, such that $\mathbf{x}_i \cdot \mathbb{S} = 0$.

Definition III.7. The *partition-productive cone* for the partition T , denoted by $\mathcal{Q}(T)$, is defined to be the cone generated by the basis vectors restricted to the stoichiometric subspace

$$\mathcal{Q}(T) = \text{cone}(\mathcal{B}_T) \perp \{\mathbf{x}_i\}.$$

As noted in the previous subsection, belonging to an FWMC partition is an equivalence relation. By definition, the species-productive cones of all equivalent autocatalytic subnetworks belonging to the partition T will lie in the partition-productive cone $\mathcal{Q}(T)$.

Example III.4. In this example, we will identify the flow-productive, species-productive and partition-productive cones for the MAS A_1 from Example III.3 (shown in Fig. 5). For $A_1 = (\{A, B, C\}, \{A + B \rightarrow C, C \rightarrow 2A\})$,

$$\begin{aligned} \bar{\mathbb{S}} &= \begin{bmatrix} -1 & 2 \\ -1 & 0 \\ 1 & -1 \end{bmatrix} \\ \bar{\mathbb{S}}_C &= \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} \\ \bar{\mathbb{S}}_C^{-1} &= \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}. \end{aligned}$$

The flow-productive cone is

$$\mathcal{F}(A_1) = \text{cone} \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right),$$

and the species-productive cone is

$$\mathcal{P}(A_1) = \text{cone} \left(\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix} \right).$$

Notice that the graph $\mathcal{G}^*$ (from Example III.2) has a positive conservation law $\mathbf{x} = [1, 1, 2]$. The partition-productive cone for the partition $T_1 = \{\{B\}, \{\}, \{\{A, C\}\}\}$ is

$$\begin{aligned} \mathcal{Q}(T_1) &= \text{cone} \left(\begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) \perp \mathbf{x} \\ &= \text{cone} \left(\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix} \right) \\ &= \mathcal{P}(A_1). \end{aligned}$$

For this example, the species-productive and partition-productive cones are identical, which is an instance of a general result proved in Theorem IV.4.

IV. ORGANIZING AN AUTOCATALYTIC ECOSYSTEM

In the previous section, we have defined and proved properties of minimal autocatalytic subnetworks

(MASs), or minimal CRNs that contain one autocatalytic core. Even mildly complicated CRNs can exhibit an abundance of autocatalytic cores [25, 27].

Definition IV.1. We define an **autocatalytic ecosystem** to be a CRN containing one or more minimal autocatalytic subnetworks (MASs).

In this section, we investigate some mathematical properties of autocatalytic ecosystems and give computational algorithms to detect and organize them. In Sec. IV A, we prove some mathematical results about MASs. In Sec. IV B and Appendix D, we provide algorithms to exhaustively enumerate the MASs and identify their species-productive cones. Finally, we discuss the polyhedral geometry of an autocatalytic ecosystem and introduce a visualization scheme in Sec. IV C.

A. Mathematical results

To understand the geometry of the different cones defined in the previous subsection, we prove some results about their behavior. First, in Proposition IV.1 we give the conditions under which two MASs with different partitions will have non-intersecting species-productive regions. Next, in Theorem IV.4, we explore conditions under which the species-productive cone is identical with the partition-productive cone for a MAS. Under such conditions, the partition itself contains the information of the species-productive regions of all the MASs in that equivalence class. Finally, in Proposition IV.5, we clarify the topological properties a CRN must possess in order for the species-productive cones of two MASs to intersect. In Sec. IV C we will use these results to construct a visualization scheme for the list of MASs in any CRN.

Proposition IV.1. *Two autocatalytic subnetworks with different food sets have disjoint species-productive regions if their non-core member species sets are empty.*

Proof. We will show that if two autocatalytic subnetworks A and B have distinct food sets and their non-core member species sets are empty, then the interiors of their species-productive cones (species-productive regions) do not intersect,

$$\text{int}(\mathcal{P}(A)) \cap \text{int}(\mathcal{P}(B)) = \emptyset.$$

Let $A_F/B_F \neq \emptyset$. Recall that the species-productive cone of the subnetworks lies within their partition-productive cone $\mathcal{Q}(\text{FWMC}(A))$ and $\mathcal{Q}(\text{FWMC}(B))$. Let $\mathbf{y}$ be the vector of length $\mathcal{S}$ such that

$$[\mathbf{y}]_i = \begin{cases} -1 & \text{if } i \in A_F/B_F \\ 0 & \text{otherwise.} \end{cases}$$

Then, by definition of the partition-productive cones, $\mathbf{y} \cdot \mathbf{v} > 0$ and $\mathbf{y} \cdot \mathbf{w} \leq 0$, where $\mathbf{v}$ and $\mathbf{w}$ are any vectors in $\text{int}(\mathcal{Q}(\text{FWMC}(A)))$ and $\text{int}(\mathcal{Q}(\text{FWMC}(B)))$, respectively. Thus, $\mathbf{y}$ defines a hyperplane separating the

two species-productive regions, and the two regions do not intersect. $\square$

Remark 21. If the non-core member species set is not empty, then the result is no longer true. For instance, see Example IV.2.

Example IV.1. Using Examples III.3 and III.4, the species-productive cones for A_1 and A_2 in the species set $\{A, B, C, D\}$ are

$$\mathcal{P}(A_1) = \text{cone} \left(\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} \right),$$

$$\mathcal{P}(A_2) = \text{cone} \left(\begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right).$$

If we let $\mathbf{y} = [0, -1, 0, 0]$, then $\mathbf{y} \cdot \text{int}(\mathcal{P}(A_1)) > 0$ and $\mathbf{y} \cdot \text{int}(\mathcal{P}(A_2)) < 0$. Thus $\mathbf{y}$ defines a separating hyperplane, and the two species-productive regions are disjoint.

Lemma IV.2. *The restriction of the species-productive cones of all MASs to their core (member) species is the non-negative orthant in the core species.*

Proof. Recall that the species-productive cone of a MAS is defined as

$$\mathcal{P}(A) = \text{cone}(\text{cols}(\bar{\mathbb{S}} \cdot \bar{\mathbb{S}}_C^{-1})).$$

By definition of the matrix inverse, the restriction of the product to the core species is the identity matrix,

$$\bar{\mathbb{S}} \cdot \bar{\mathbb{S}}_C^{-1}|_C = I_C.$$

Thus, the species-productive cone restricted to the core species is the non-negative orthant. $\square$

As remarked in Remark 16, a MAS cannot contain an inflow or outflow reaction. Moreover, for material CRNs, where the species are made of constituents that are conserved in a reaction, a MAS must also possess at least one conservation law. The following results are applicable for such MASs. (In Sec. V, we develop the formalism for CRNs with non-negative integer conservation laws and find their application in the examples.)

Lemma IV.3. *If a CRN has any positive conservation laws, the union of the food set and the non-core member set of any MAS must be non-empty.*

Proof. Let A be a MAS, with associated stoichiometric matrix $\bar{\mathbb{S}}$, of a CRN with a positive conservation law given by the vector $\mathbf{x}$. Assume for the sake of contradiction that A does not have any element in the food set A_F or the non-core member set $A_{M/C}$. For any flow $\mathbf{v}$ in the productive region of A , we know that $\bar{\mathbb{S}} \cdot \mathbf{v} \gg 0$. Also,

since $\mathbf{x}$ is a positive conservation law, $\mathbf{x} \cdot \bar{\mathbb{S}} \cdot \mathbf{v} \stackrel{!}{=} 0$. Since the sum of positive values cannot add up to zero, we have a contradiction. Thus, there must be at least one element in the food set or the non-core member set of a MAS with positive conservation laws. (See example IV.2.) $\square$

Theorem IV.4. *For a CRN with positive conservation laws, the species-productive cone of any MAS with exactly one more species than the core set is identical to its partition-productive cone.*

Proof. Let A be a MAS of a CRN with positive conservation laws $\{\mathbf{x}_i\}$ with exactly one more species than the core set, denoted by $C = \{C_1, \dots, C_C\}$. From Lemma IV.3, we know that the extra species must be either in the food set or non-core member set of A . In either case, the species must be net consumed by the subnetwork to respect the positive conservation laws, and we denote it by f .

Let us label the species set as $\{f, C_1, \dots, C_C\}$. Recall that the partition-productive cone is defined as $\mathcal{Q}(\text{FWMC}(A)) = \text{cone}(\mathcal{B}) \perp \{\mathbf{x}_i\}$, where

$$\mathcal{B} = \left\{ \begin{bmatrix} -1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\}.$$

Note that the additional basis vector for the non-core member species with positive entries is omitted since we know that it must be consumed in the species-productive region to respect the positive conservation laws. Moreover, using Lemma IV.2, we know that the restriction of the species-productive cone to the core species is the non-negative orthant in the core species. In particular, for the species-productive cone $\mathcal{P}(A) = \text{cone}(\text{cols}(\bar{\mathbb{S}} \cdot \bar{\mathbb{S}}_C^{-1}))$, we have

$$\bar{\mathbb{S}} \cdot \bar{\mathbb{S}}_C^{-1} = \begin{bmatrix} \mathbf{f} \\ I_C \end{bmatrix},$$

where $\mathbf{f}$ is a row vector $[\mathbf{f}_1, \dots, \mathbf{f}_C]$ and I_C is the identity matrix of size C . Since the CRN also has conservation laws, these coefficients must satisfy

$$\begin{aligned} [\mathbf{x}_i]_f \cdot \mathbf{f}_1 + [\mathbf{x}_i]_{C_1} \cdot 1 &= 0 \\ &\vdots \\ [\mathbf{x}_i]_f \cdot \mathbf{f}_C + [\mathbf{x}_i]_{C_C} \cdot 1 &= 0, \end{aligned}$$

for every conservation law indexed by i . But these are simply the basis vectors of the partition-productive cone $\mathcal{B}$ when made orthogonal to the conservation laws. Thus the partition-productive cone is identical to the species-productive cone. $\square$

Example IV.2. Consider two MASs (from Sec. V B 1 Fig. 3)

$$\begin{aligned} A_1 &= (\{\bar{1}, \bar{2}, \bar{3}, \bar{4}\}, \{\bar{1} + \bar{3} \rightarrow \bar{4}, \\ &\quad \bar{1} + \bar{4} \rightarrow \bar{2} + \bar{3}, \bar{2} + \bar{2} \rightarrow \bar{3} + \bar{1}\}), \\ A_2 &= (\{\bar{1}, \bar{2}, \bar{3}, \bar{4}\}, \{\bar{1} + \bar{3} \rightarrow \bar{4}, \\ &\quad \bar{1} + \bar{4} \rightarrow \bar{2} + \bar{3}, \bar{2} + \bar{2} \rightarrow \bar{4}\}). \end{aligned}$$

The partitions of A_1 and A_2 are $\{\{\}, \{\}, \{\bar{1}, \{\bar{2}, \bar{3}, \bar{4}\}\}\}$ and $\{\{\bar{1}\}, \{\}, \{\{\bar{2}, \bar{3}, \bar{4}\}\}\}$, respectively. A simple calculation yields that their species-productive cones are in fact identical, and

$$\mathcal{P}(A_1) = \mathcal{P}(A_2) = \text{cone} \left(\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right).$$

Definition IV.2. For a CRN $\mathcal{G} = (\mathcal{S}, \mathcal{R})$, the reversible CRN $\mathcal{G}_{\text{rev}} = (\mathcal{S}, \mathcal{R}')$ is defined such that for every reaction $y_1 \rightarrow y_2 \in \mathcal{R}$, the set $\mathcal{R}'$ contains both $y_1 \rightarrow y_2$ and $y_2 \rightarrow y_1$.

Remark 22. Reversible CRNs are worth considering insofar as chemistry often assumes that all reactions are theoretically reversible (even if the rate constants of forward and reverse reactions differ greatly in magnitude).

Proposition IV.5. *If the species-productive cones of two MASs have an intersection, the reversible CRN of their union has a null flow.*

Proof. Let the productive cones of two MASs A and B , have a non-empty intersection, $\mathcal{P}(A) \cap \mathcal{P}(B) \neq \emptyset$. Let $A \cup B$ be the subnetwork obtained by taking the union of the two subnetworks, and let $\mathbb{S}_{A \cup B}$ be the associated stoichiometric matrix. Then, there must be non-identical flows $\mathbf{v}_A$ and $\mathbf{v}_B$ with support in A and B , respectively, such that

$$\mathbb{S}_{A \cup B} \cdot \mathbf{v}_A = \mathbb{S}_{A \cup B} \cdot \mathbf{v}_B.$$

This implies that $\mathbb{S}_{A \cup B} \cdot (\mathbf{v}_A - \mathbf{v}_B) = 0$, and thus the kernel of the stoichiometric matrix is non-empty. In general, the difference of their flows $(\mathbf{v}_A - \mathbf{v}_B)$ can have negative entries, however, it yields the interpretation of a null flow on a network if the union of the two subnetworks $A \cup B$ is made reversible. $\square$

Remark 23. The converse statement, if the reversible CRN of the union of two MASs contains a null cycle then their species-productive cones must intersect, is not true. For example, consider the MASs (from Sec. V B 1 Fig. 3)

$$\begin{aligned} A_1 &= (\{\bar{1}, \bar{2}, \bar{3}\}, \{\bar{1} + \bar{2} \rightarrow \bar{3}, \bar{1} + \bar{3} \rightarrow \bar{2} + \bar{2}\}), \\ A_2 &= (\{\bar{1}, \bar{2}, \bar{3}\}, \{\bar{2} \rightarrow \bar{1} + \bar{1}, \bar{3} + \bar{1} \rightarrow \bar{2} + \bar{2}\}). \end{aligned}$$

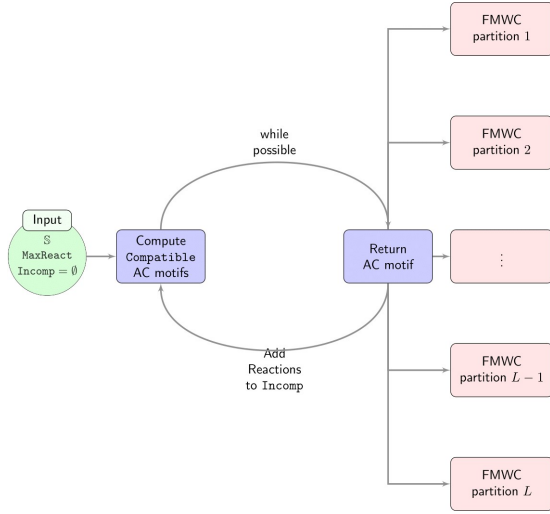


Figure 6: Flowchart of our procedure to construct minimal autocatalytic subnetworks.

Notice that $\text{FWMC}(A_1) = \{\{\bar{1}\}, \{\}, \{\{\bar{2}, \bar{3}\}\}\}$ and $\text{FWMC}(A_2) = \{\{\bar{3}\}, \{\}, \{\{\bar{1}, \bar{2}\}\}\}$. Thus, using Proposition IV.1, the two species-productive cones do not intersect. However, the union of the two subnetworks yields

$$A_1 \cup A_2 = (\{\bar{1}, \bar{2}, \bar{3}\}, \{\bar{1} + \bar{2} \rightarrow \bar{3}, \bar{1} + \bar{3} \rightarrow \bar{2} + \bar{2}, \bar{2} \rightarrow \bar{1} + \bar{1}\}).$$

which has a null cycle (with null flow $[1, 1, 1]^T$).

B. Algorithms

In what follows we describe a mathematical optimization based approach to enumerate the whole list of exclusively autocatalytic cores in a CRN. The model that we propose uses the following variables:

$$y_i = \begin{cases} 1 & \text{if species } i \in \text{Core set,} \\ 0 & \text{otherwise} \end{cases}.$$

for all $i \in \mathcal{S}$.

$\mathbf{v} = [v_1, \dots, v_{|\mathcal{R}|}] \in [0, 1]_+^{|\mathcal{R}|}$: flow inducing a MAS.

$$z_r = \begin{cases} 1 & \text{if } v_r > 0, \\ 0 & \text{otherwise} \end{cases}, \text{ for all } r \in \mathcal{R}.$$

These variables are adequately combined by means of linear constraints to ensure the verification of the properties of Theorem III.1. Specifically:

- The species in the Core set ($y_i = 1$) are the productive species of the selected reactions:

$$\sum_{r \in \mathcal{R}} (\mathbb{S})_i^r v_r \geq \epsilon - \Delta_i(1 - y_i), \forall i \in \mathcal{S}, \quad (8)$$

Note that in case $y_i = 1$, the linear inequality is activated ensuring that the i -th species is produced in net. Otherwise, the constraint is redundant.

- All species in the Core set are both consumed and produced by the reactions in the core:

$$\sum_{\substack{r \in \mathcal{R}: \\ (\mathbb{S}^+)_i^r > 0}} z_r \geq y_i, \quad \forall i \in \mathcal{S}, \quad (9)$$

$$\sum_{\substack{r \in \mathcal{R}: \\ (\mathbb{S}^-)_i^r > 0}} z_r \geq y_i, \quad \forall i \in \mathcal{S}, \quad (10)$$

The two sets of inequalities above ensure that in case $y_i = 1$, there must exist reactions in the core (those with $z_r = 1$) for which species i with $(\mathbb{S}^+)_i^r > 0$ (for (9)) and $(\mathbb{S}^-)_i^r > 0$ (for (10)).

- All reactions in the core both consume and produce core species:

$$\sum_{\substack{i \in \mathcal{S}: \\ (\mathbb{S}^+)_i^r > 0}} y_i \geq z_r, \quad \forall r \in \mathcal{R}, \quad (11)$$

$$\sum_{\substack{i \in \mathcal{S}: \\ (\mathbb{S}^-)_i^r > 0}} y_i \geq z_r, \quad \forall r \in \mathcal{R}. \quad (12)$$

In case reaction r is part of the core ($z_r = 1$), there exists at least one produced core species (inequality (11)) and at least one consumed species (inequality (12)).

- The non-selected reactions have zero-flow:

$$v_r \leq z_r, \forall r \in \mathcal{R}. \quad (13)$$

In case reaction r is not included in the core ($z_r = 0$), the flow component for this reaction is forced to be zero ($v_r = 0$). Otherwise, the flow is unrestricted ($v_r \leq 1$).

Then, the master mixed integer programming (MILP) problem that allows us to generate the entire list of exclusively autocatalytic cores consists of minimizing the number of reactions in a network with all the above conditions:

$$\begin{aligned} & \min \sum_{r \in \mathcal{R}} z_r \\ & \text{s.t. (8) - (13),} \\ & y_i \in \{0, 1\}, \forall i \in \mathcal{S}, \\ & z_r \in \{0, 1\}, \forall r \in \mathcal{R}, \\ & \mathbf{v} \in [0, 1]_+^{|\mathcal{R}|}. \end{aligned}$$

The general procedure consists of sequentially solving the above optimization problem and adding to it new linear constraints to allow the generation of different cores. A flowchart of our procedure to construct the minimal exclusively autocatalytic cores is shown in Fig. 6. Define a list of *sets of incompatible reactions*, **Incomp**, initialized to the empty set, that stores, at each iteration, the sets of reactions that are part of the already computed MASs,

and avoids generating MASs containing those sets of reactions. Then, in the first step, our algorithm computes an autocatalytic subnetwork by selecting a set of core species and reactions taking part in a subnetwork with minimum number of reactions. The obtained subnetwork is stored, and its FWMC partition is computed. This set of reactions, say $\mathcal{R}'$, is added to **Incomp**, and the procedure is repeated, by adding the following constraint to the MILP to assure that the subnetworks obtained in future iterations do not contain any of the sets of reactions in **Incomp**:

$$\sum_{r \in \mathcal{R}'} z_r \leq |\mathcal{R}'| - 1.$$

The algorithm stops when there does not exist sets of reactions not *incompatible* with those in **Incomp**. Minimality is assured by the minimization of the number of reactions in the subnetwork in each run and by the conditions imposed by the set **Incomp**. Note that the complete enumeration of all the possible combinations of reactions/species that may be included in a MAS, and the detection of a MAS is computationally prohibitive in practice. However, our procedure avoids this enumeration by solving an MILP at each iteration. Moreover, if one is interested on finding just a certain number, k , of MAS, it can be done by terminating the algorithm after k iterations and giving as output the k obtained MASs.

The identification of the FWMC partition for a given autocatalytic subnetwork results from checking both the signs of the restricted stoichiometric matrix, $\bar{S}$ and the production vector $\bar{S}v^*$ for the flow v^* also obtained after each iteration. Specifically, for each species i in the subnetwork: if all the components in $\bar{S}_i$ are non positive (resp. non negative), i is classified as a food (resp. waste) species; in case $\bar{S}_i$ has negative and positive entries, i is classified as a member species; and if $\bar{S}_i$ has negative and positive entries and $(\bar{S}v^*)_i$ is strictly positive, then i is a core species.

Once the set of MASs is obtained, they are classified by means of their FWMC partitions. (Additional checks must be made to determine if there is more than one autocatalytic core in a MAS. In this case, the FWMC partition of the subnetwork is non-unique, and the MAS must be assigned to all the FWMC classes to which its autocatalytic cores belong.) Within each FWMC class, we check the *pairwise intersection* of the species-productive cones. For each subnetwork A_1, A_2 in the same class, we first check whether the cones $\mathcal{P}(A_1)$ and $\mathcal{P}(A_2)$ intersect by finding a nonzero vector shared by both cones. In case the cones intersect, we check whether the cones are identical, one is contained in the other, or they partially intersect. This check is performed by computing the distances between the (normalized) generators of the cones. If all these distances are zero, the cones are identical; if one of the sets of distances is zero but the other is not, one of the cones is contained in the other; but if in both sets there are positive distances, the cones partially intersect. Using the same procedure, we also check the

pairwise intersection of the partition-productive cones of each equivalence class. Further details on the described approaches are provided in Appendix D 1.

C. Geometry and visualization

In the last subsection, we proposed an algorithm that, given a stoichiometric matrix of a CRN, outputs a list of the MASs it contains. In this section, we will explain how to take the list of MASs and visualize their combinatorics and geometry. Recall that for each MAS, we define a flow-productive cone in the flow space of the graph where each core member species is strictly produced, and a species-productive cone on the space of changes in concentration (population). Geometrically, the space of changes in chemical concentration (i.e., the velocity space of chemical concentrations) is the stoichiometric subspace of the CRN. Thus, the list of MASs can be seen as yielding a partial polyhedral decomposition of the stoichiometric space induced from the flows on the hypergraph (CRN). We remark that, for a more complete decomposition one must also consider the consumptive cones of the minimal drainable subnetworks. However, it is not necessary that the union of the autocatalytic and drainable subnetworks span the complete stoichiometric space (for example, see Sec. VB 2 Fig. 14).

In subsec. IIIB, we showed that each subnetwork can be assigned its FWMC equivalence class(es). As explained in Proposition IV.1, while there are cases where the species-productive cones of different equivalence classes can be shown to not intersect, in general the productive regions of different equivalence classes can intersect. To visualize the list of equivalence classes to which the MASs belong, we run the algorithm for finding an intersection between each pair of partition-productive cones and obtain a two-dimensional square matrix, C_{pp} , of dimension equal to the number of equivalence classes. Let us denote the list of equivalence classes by $\mathcal{L}$. The entry of C_{pp} in the i^{th} row and j^{th} columns is given by,

$$[C_{pp}]_i^j = \begin{cases} 2 & \text{if } Q(\mathcal{L}_i) = Q(\mathcal{L}_j), \\ 1 & \text{if } Q(\mathcal{L}_i) \subsetneq Q(\mathcal{L}_j), \\ 0 & \text{if } \emptyset \neq Q(\mathcal{L}_i) \cap Q(\mathcal{L}_j) \subsetneq Q(\mathcal{L}_i), Q(\mathcal{L}_j), \\ -1 & \text{if } Q(\mathcal{L}_i) \cup Q(\mathcal{L}_j) = \emptyset. \end{cases}$$

Notice that any asymmetry in entries across the diagonal indicates that only one of the partition-productive cones completely contain the other. This matrix can then be visualized as a heat map, for example see Fig. 7.

Each equivalence class can contain several MASs. The species-productive cones of the MASs in a class are not always identical. While they will share the same projection on the core species, the productive regions in the non-core species can be very different. For example, let us pick the equivalence class $\{\{4\}, \{1\}, \{\{2, 3, 5\}\}\}$ from the list of classes shown in Fig. 7. From the figure, we know that it contains 5 MASs. We plot the projection

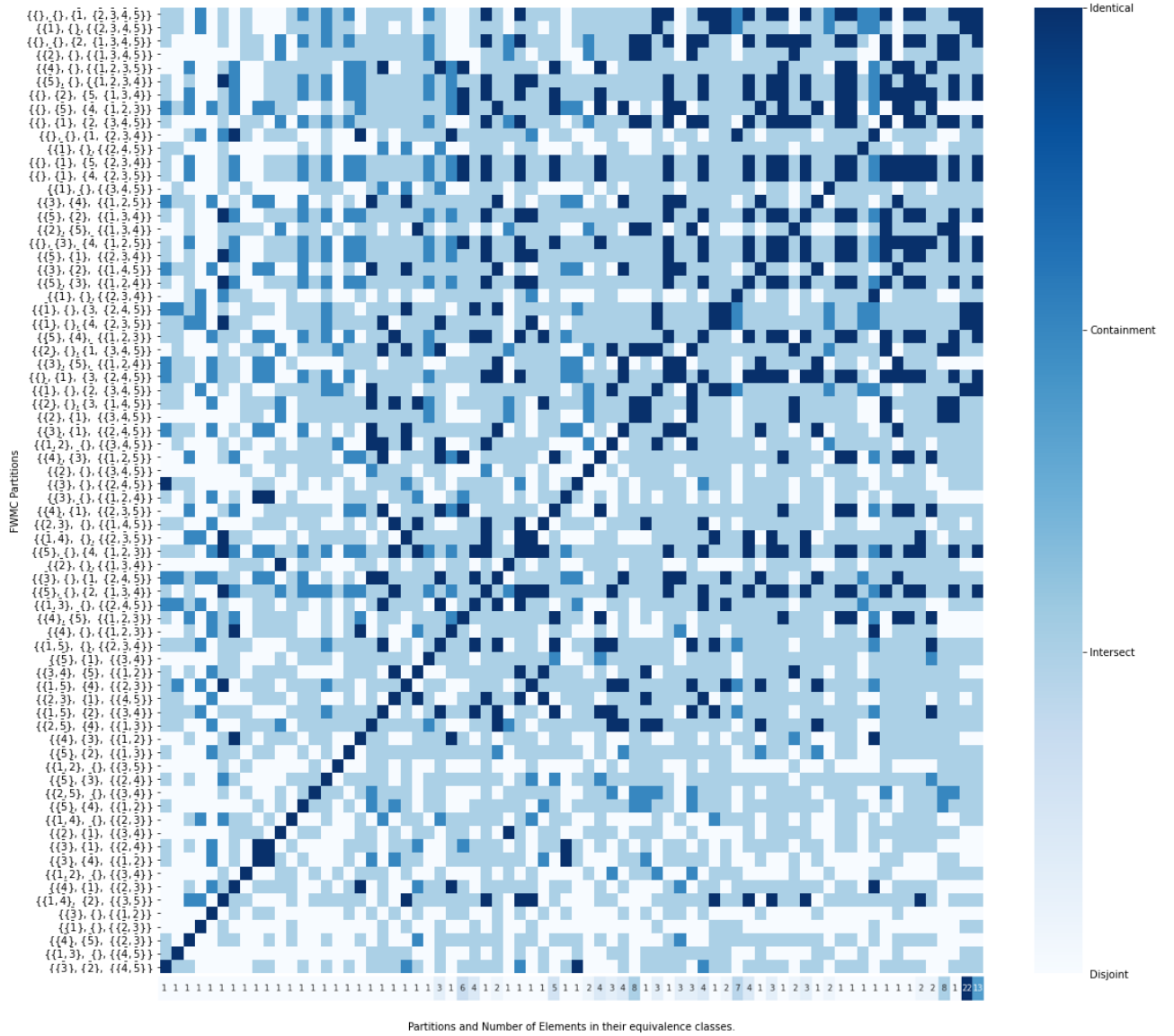


Figure 7: The intersection data of partition-productive cones for all partitions of the complete $L = 5$ 1-constituent CCRN (described in Sec. VB 1).

of the species-productive region in the non-core species in the top panel of Fig. 8. In higher dimensions when more non-core members are involved, this representation can get rather cumbersome. Thus, we employ a similar visualization as C_{pp} to depict the intersection of MASS within an equivalence class. Whether or not there is an intersection between the species-productive cones can be ascertained using our algorithm outlined in the previous subsection. For an example of the resulting visualization for the same equivalence class considered above, see the bottom panel of Fig. 8. In the same manner, a visualization for the information of pairwise intersection of the productive cones of all MASSs for a CRN can also be obtained, for example see Fig. 9.

Example IV.3. Consider the complete $L = 5$ 1-

constituent CCRN of order two (introduced in Section V). The list of all the FWMC classes, the number of MASs they contain, and their intersection information is summarized in Fig. 7. Consider the class $\{\{4\}, \{\bar{1}\}, \{\{2, \bar{3}, \bar{5}\}\}\}$. It contains the MASs:

$$\begin{aligned} \text{AS1} &= \{\bar{2} + \bar{2} \rightarrow \bar{3} + \bar{1}, \\ &\quad \bar{5} \rightarrow \bar{2} + \bar{3}, \\ &\quad \bar{4} + \bar{3} \rightarrow \bar{2} + \bar{5}\}, \\ \text{AS2} &= \{\bar{4} + \bar{2} \rightarrow \bar{1} + \bar{5}, \\ &\quad \bar{5} \rightarrow \bar{2} + \bar{3}, \\ &\quad \bar{4} + \bar{3} \rightarrow \bar{2} + \bar{5}\}, \end{aligned}$$

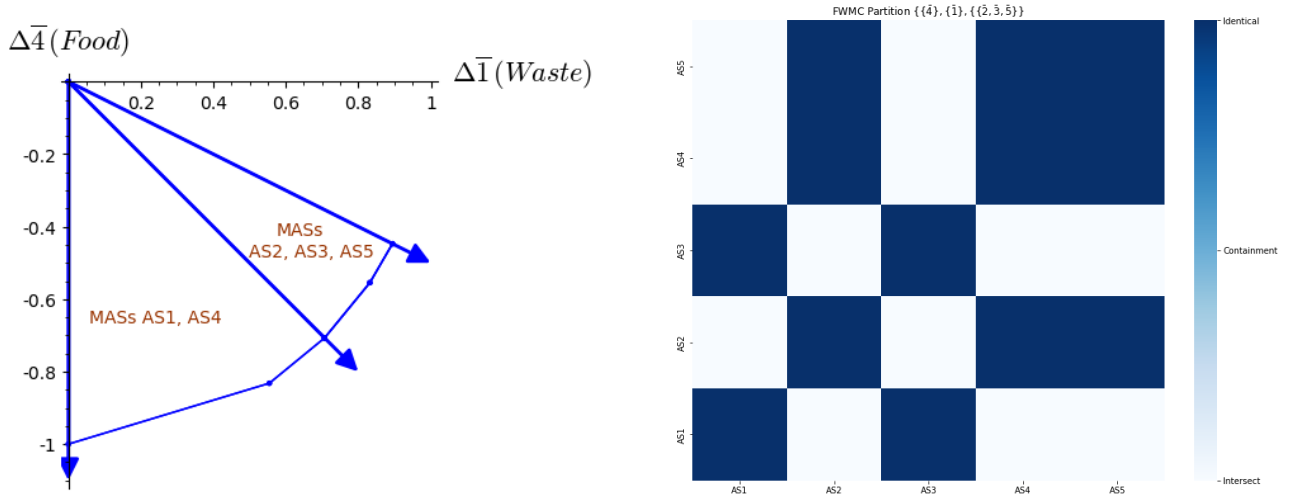
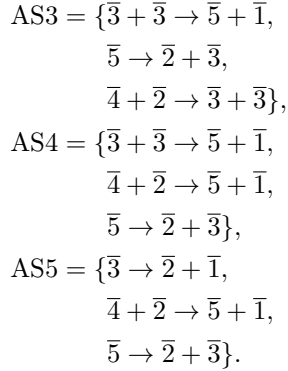


Figure 8: All MASs belonging to the $\{\{\bar{4}\}, \{\bar{1}\}, \{\{\bar{2}, \bar{3}, \bar{5}\}\}\}$ class in the complete $L = 5$ 1-constituent CCRN are considered. The projections of their species-productive cones on the food and waste species are shown in the left panel (where the change in the concentration of species $\bar{n}$ is denoted as $\Delta \bar{n}$). The right panel displays the intersection information of MASs within a class analogous to Fig. 7. Note that there are no MASs in the class that are disjoint, and thus the ‘disjoint’ label is omitted.



The projection of the species-productive cones for the above MASs and their intersection information is summarized in Fig. 8.

V. CLUSTER CHEMICAL REACTION NETWORKS (CCRN) FRAMEWORK

All fully-balanced chemical reactions conserve the number of atoms of each type. Often, but not always, reaction models of polymerization and hydrolysis also conserve the number of each monomer. The existence of such conservation laws allows us to coarse-grain the underlying detailed CRNs, projecting away the bond structure of each molecule or polymer and only counting the number of atoms or monomers, respectively. Following terminology from Chapter 7 of [28], we will refer to this coarse-grained description of each species as a *cluster* (see Fig. 10). This also allows us to coarse-grain a realistic CRN into a new CRN which has clusters as species and reactions between multisets of clusters induced from the

reactions in the original CRN. We will call the resulting CRN, a *cluster chemical reaction network* (CCRN) obtained from the original CRN.

There are three advantages of considering CCRN descriptions of CRNs. First, consider a CRN whose species set consists of all linear polymers formed of D monomers up to length N . Then, the CCRN induced by the CRN would reduce the number of species from exponential (D^N) to polynomial (N^D) in the length of the polymer, reducing number of species and changing the connectivity in the reaction graph from a tree to a lattice for many reaction models. Secondly, since each reaction in the original CRN induces a reaction in the CCRN with the same topology, the mapping would preserve the autocatalytic property. In other words, if a CRN is autocatalytic, its induced CCRN will also be autocatalytic. However, since it is a many-to-one map, multiple autocatalytic subnetworks of the CRN might map to the same autocatalytic subnetwork in its CCRN or the CCRN might contain MASs that do not correspond to any motif in the CRN (see Example V.1). Thirdly, one can later re-introduce structure into the CCRN by adding more species with the same cluster counts (conserved quantities). In this way, the coarse-graining can be systematically refined to recover the original CRN by adding new dimensions, and allows a gradual complexification of the model.

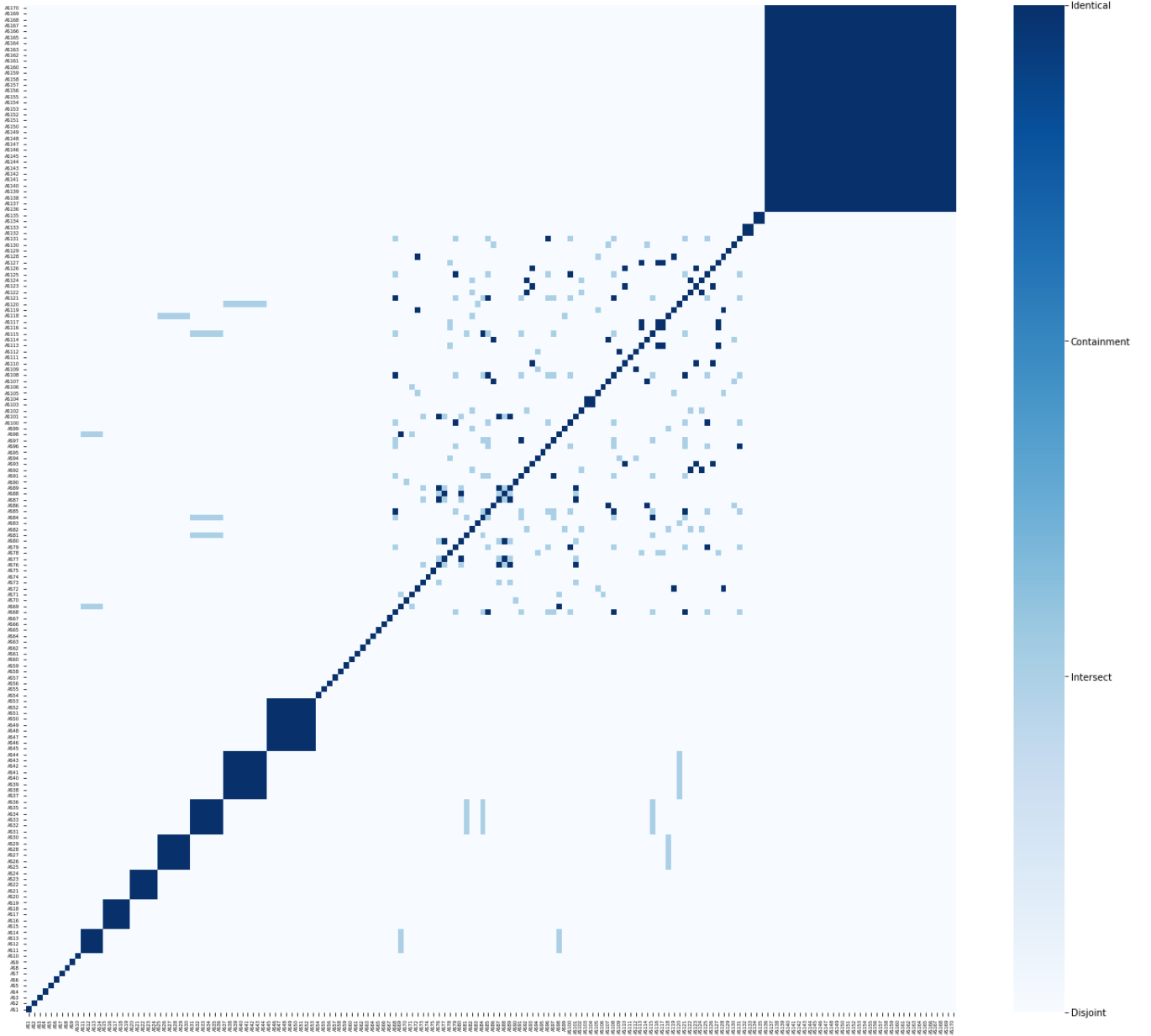


Figure 9: The pairwise intersection statistic for all the species-productive cones of the $L = 5$ 1-constituent complete CCRN of order two. The reaction network and the list of MASs can be found in Appendix E.

Example V.1. Consider the CRN

$$\begin{aligned}
 \mathcal{G} &= (\mathcal{S}, \mathcal{R}) \\
 \mathcal{S} &= \{A, B, BAA, ABB, ABBA, BAAB, ABBABAAB\} \\
 \mathcal{R} &= \{ABBA + BAA + B \rightarrow ABBABAAB, \\
 &\quad BAAB + ABB + A \rightarrow ABBABAAB, \\
 &\quad ABBABAAB \rightarrow ABBA + BAAB.\}
 \end{aligned}$$

The CCRN obtained by counting the number of As and Bs in each string and representing the resulting clusters

as $\overline{n_A, n_B}$ is

$$\begin{aligned}
 \text{CCRN}(\mathcal{G}) &= (\mathcal{S}_C, \mathcal{R}_C) \\
 \mathcal{S}_C &= \{\overline{1, 0}, \overline{0, 1}, \overline{1, 1}, \overline{2, 2}, \overline{2, 4}, \overline{4, 4}\} \\
 \mathcal{R}_C &= \{\overline{2, 2} + \overline{2, 1} + \overline{0, 1} \rightarrow \overline{4, 4}, \\
 &\quad \overline{2, 2} + \overline{1, 2} + \overline{1, 0} \rightarrow \overline{4, 4}, \\
 &\quad \overline{4, 4} \rightarrow \overline{2, 2} + \overline{2, 2}\}
 \end{aligned}$$

Notice that while the CRN has one minimal autocatalytic subnetwork (MAS), the CCRN has two MASs. This can be remedied by remembering that $BAAB$ and $ABBA$ are distinct by introducing a new species $\overline{2, 2}^*$. Resulting

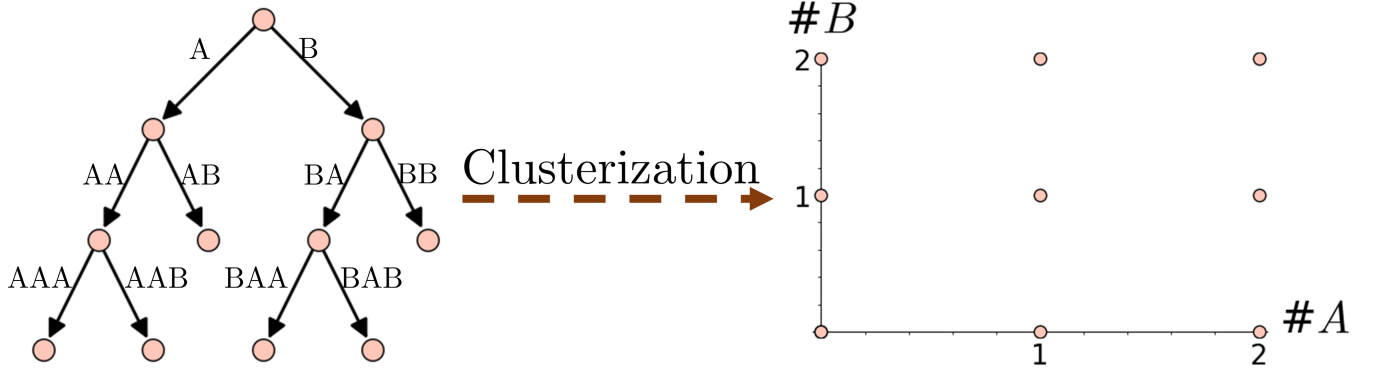


Figure 10: Consider a CRN of polymers whose species set is arranged as a binary tree in two letters, shown to the left. Then the *cluster* resulting from each binary string is obtained by counting the number of occurrences of each alphabet. The *clusterization* process projects from a tree structure over polymers to a lattice structure over clusters, shown to the right.

in the CCRN

$$\begin{aligned} \text{CCRN}(\mathcal{G})^* &= (\mathcal{S}_C^*, \mathcal{R}_C^*) \\ \mathcal{S}_C^* &= \{\overline{1}, \overline{0}, \overline{0}, \overline{1}, \overline{2}, \overline{1}, \overline{1}, \overline{2}, \overline{2}, \overline{2}, \overline{2}^*, \overline{4}, \overline{4}\} \\ \mathcal{R}_C^* &= \{\overline{2}, \overline{2} + \overline{2}, \overline{1} + \overline{0}, \overline{1} \rightarrow \overline{4}, \overline{4}, \\ &\quad \overline{2}, \overline{2}^* + \overline{1}, \overline{2} + \overline{1}, \overline{0} \rightarrow \overline{4}, \overline{4}, \\ &\quad \overline{4}, \overline{4} \rightarrow \overline{2}, \overline{2} + \overline{2}, \overline{2}^*.\} \end{aligned}$$

Note that $\text{CCRN}(\mathcal{G})^*$ has identical information to the original CRN and should, therefore, yield identical inferences to the original CRN.

We formally define a CCRN in Sec. V A, and systematically explore the properties of CCRNs with one conserved quantity (type of atom or monomer) in Sec. V B. We also provide the statistics of autocatalytic subnetworks found in such reaction networks and present a worked-out-example in Sec. V B 2. We briefly discuss the computational challenges in scaling the algorithm for fully connected models in Sec. V B 4 and introduce rule-generated CCRNs in Sec. V C.

A. Formalism

A **cluster chemical reaction network (CCRN)** is a CRN with the additional structure that, upon excluding the inflow and outflow reactions (see Remark 16), the network has at least one nonnegative integer conservation law (see Sec. II D). Each conservation law $\mathbf{x}$ yields a conserved quantity $\mathbf{x} \cdot \mathbf{n}$ for a population in state $\mathbf{n}$. Each conservation law corresponds to a type of **constituent** that is conserved, and the magnitude of the conserved quantity denotes the amount of that constituent in the population state. We denote the number of distinct constituents by D and label them as $A_1, \dots, A_D$. The *species* of a CCRN, termed **clusters**, are multisets of constituents and denoted by an overline (notation chosen

to be consistent with [29]) over a vector of nonnegative integers representing the number of each constituent in the cluster.

Notation V.1.

$$\overline{\mathbf{n}} := \overline{n_1, \dots, n_D}$$

will be used to denote a cluster comprising $n_1, \dots, n_D$ of constituents $A_1, \dots, A_D$, respectively.

Denoting the population state with exactly one $\overline{\mathbf{n}}$ type particle as $1_{\overline{\mathbf{n}}}$, if the conservation laws are labelled $\mathbf{x}_d$ where d takes values from 1 to D , then

$$\mathbf{x}_d \cdot 1_{\overline{\mathbf{n}}} = n_d.$$

A cluster thus corresponds to the vector of its conserved quantities. In chemistry, the compositional formula of a chemical is an example of its cluster representation. For example, the cluster of water or H_2O , in a basis of constituents $\{H, O\}$ will be represented as $\overline{2}, \overline{1}$. Here the first and second conservation laws counts the number of H and O atoms, respectively. In case of multiple clusters with identical vectors of conserved quantities, we distinguish them with an asterisk *. For example, in Table II, $\overline{1}$ and $\overline{1}^*$ refer to a monomer and an activated monomer, respectively.

Definition V.1. We define the **length** of a cluster to be the sum of its conserved quantities,

$$\ell(\mathbf{n}) := \sum_{i=1}^D n_i.$$

The length of a cluster is a scalar quantity.

In a CCRN, **complexes** are multisets of clusters and are denoted by

$$c^\alpha := \sum_{\mathbf{n}} c_{\mathbf{n}}^\alpha \overline{\mathbf{n}}.$$

As an abuse of notation, we denote the stoichiometry of complex c^α also by c^α , where now it is the column vector with entries c_n^α .

Definition V.2. The **size per constituent** of a complex is the vector sum of conserved quantities of each type, and is denoted

$$|c^\alpha| = \sum_{\mathbf{n}} c_{\mathbf{n}}^\alpha \mathbf{n}.$$

Definition V.3. The **width** of a complex is the total number of clusters in the multiset, and denoted by

$$w(c^\alpha) := \sum_{\mathbf{n}} c_{\mathbf{n}}^\alpha.$$

In a CCRN, every **reaction** must be such that the sizes per constituent of the source and target complexes are identical. Thus, a reaction $c^\alpha \rightarrow c^\beta$ is allowed only if

$$|c^\alpha| = \sum_{\mathbf{n}} c_{\mathbf{n}}^\alpha \mathbf{n} = \sum_{\mathbf{n}} c_{\mathbf{n}}^\beta \mathbf{n} = |c^\beta|.$$

This restriction is required in order for the constituents to be conserved quantities, as

$$\sum_{\mathbf{n}} \mathbf{n}(c^\beta - c^\alpha)_{\mathbf{n}} = 0.$$

Definition V.4. Let $\mathcal{H}_D = (\mathcal{S}, \mathcal{R})$ denote a CCRN with D constituents. Then for any finite CCRN,

$$\begin{aligned} \mathcal{S} &\subset \{\bar{\mathbf{n}} | \mathbf{n} \in \mathbb{Z}_{\geq 0}^D\}, \\ \mathcal{R} &\subseteq \{c^\alpha \rightarrow c^\beta | |c^\alpha| = |c^\beta|\}. \end{aligned}$$

For a CCRN $\mathcal{H}_D$:

- The **length** of the CCRN is defined to be the maximum length of its clusters, and denoted

$$\ell(\mathcal{H}_D) = \max_{\mathbf{n} \in \mathcal{S}} \ell(\mathbf{n}).$$

- The **order** of the CCRN is defined to be the maximum width of its complexes, and denoted by

$$o(\mathcal{H}_D) = \max_{\mathcal{R}} w(c^\alpha).$$

Definition V.5. We define a **complete CCRN** of length L and order w to consist of all species up to length L and all possible reactions that are allowed by the conservation laws with complexes of width up to w .

Remark 24. Since any reaction consisting of more than two reactants or products can be written as chains of second-order reactions, for our analysis *we will restrict to CCRNs of order two*. Notice that this restriction ensures that any exclusively autocatalytic subnetwork will also be stoichiometrically autocatalytic because the reactants and products are necessarily distinct.

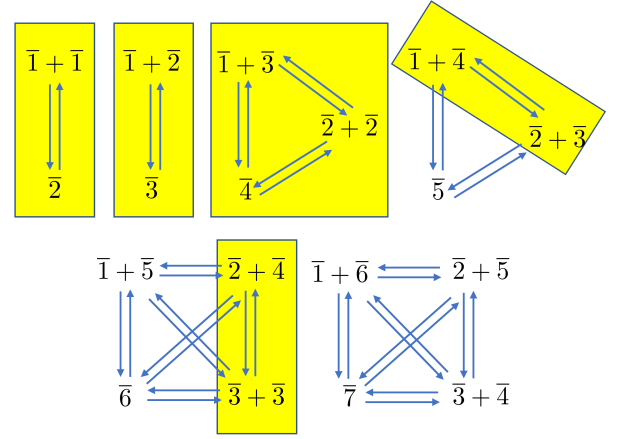


Figure 11: The first few reactions in the complete 1-constituent CCRN of order two. The reactions containing clusters only up to length four ($L = 4$) are shown in yellow boxes.

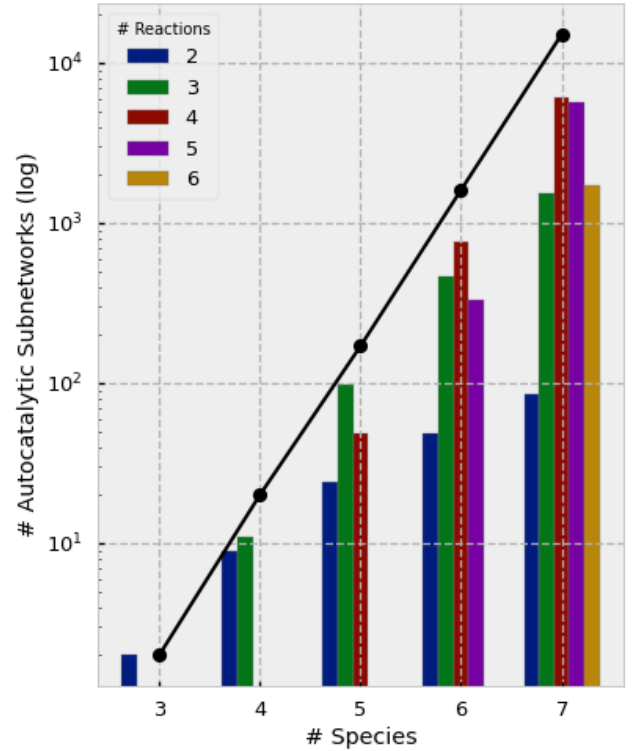


Figure 12: Number of minimal autocatalytic subnetworks (MASs) varying with L for the complete 1-constituent CCRN of order two. For each L , the distribution of MAS with the number of reactions is also shown. Notice that total MASs for each L , plotted as black dots, increases exponentially with L .

Max. length L	CCRN properties					# reactions in a MAS				
	$ \mathcal{C} $	$ \mathcal{R} $	ℓ	ι	δ	2	3	4	5	6
3	6	6	3	3	1	2	0	0	0	0
4	11	14	5	8	3	9	11	0	0	0
5	17	26	7	16	6	24	98	48	0	0
6	24	44	9	29	10	48	461	768	331	0
7	32	68	11	47	15	85	1549	6028	5673	1709

Table I: The complete 1-constituent CCRN of length L and order two is considered for L from 3 – 7. The topological properties of the CCRN, such as number of linkage classes ℓ and deficiency δ , are shown in the left half (for calculations, see Appendix B). The number of MASs for a given number of reactions is shown in the right half (for a visual representation, see Fig. 12).

B. Complete CCRN

1. 1-constituent CCRN

Let $\mathcal{H}_1^L = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ denote a complete 1-constituent CCRN of length L and order two. 1-constituent CCRN of length L means that the species set consists of clusters up to length L , denoted $\mathcal{S} = \{\bar{1}, \bar{2}, \dots, \bar{L}\}$. By order two, we mean that all complexes are at most of width two, and the set of complexes is denoted as $\mathcal{C} = \{\bar{a}\} \cup \{\bar{a} + \bar{b}\}$ for $\bar{a}, \bar{b} \in \mathcal{S}$ and $b \leq a \leq L$. By a *complete CCRN*, we mean that the set of reactions contains all the possible reactions that are allowed by the conservation law, i.e.

$$\mathcal{R} = \{\bar{a} + \bar{b} \rightleftharpoons \bar{c} + \bar{d} | a + b = c + d\} \cup \{\bar{a} \rightleftharpoons \bar{c} + \bar{d} | a = c + d\}.$$

The first few linkage classes of the complete 1-constituent CCRN of order two are shown in Fig. 11 and the complete CCRN for $L = 4$ is highlighted in yellow. The information about the reaction networks and the MASs that they contain for the complete 1-constituent CCRNs of L ranging from 3 to 7 is collected in Table I and Fig. 12.

Consider the complete 1-constituent CCRN with $L = 4$ of order two, the reaction network for which is shown in Fig. 11. Notice that this subnetwork has 5 linkage classes, 11 complexes, and 14 reactions. An application of the algorithm for enumerating all the MASs of the CCRN discussed in Sec. IV B yields a list of 20 subnetworks, as shown in Fig. 3. In the list, there are 9 MASs with two reactions and 11 MASs with three reactions. Since the CCRN has 4 species, and 1 conservation law, the stoichiometric subspace is of dimension 3. Moreover, due to the conservation law, from Lemma IV.3 each MAS must possess at least one non-core species. This means that there cannot be any 4 reaction MAS as it would require at least 5 distinct species, which would be a contradiction.

Upon obtaining a list of MASs for a given CRN, we would like to understand how their species-productive

and partition-productive cones intersect. Observe that each MAS in the $L = 4$ CCRN considered above has a unique autocatalytic core. For the CCRN, using Theorem IV.4, any MASs with the same FWMC partitions have identical species-productive cones and it is identical to the partition-productive cone. Thus, we do not show any plot for the species-productive cone intersection data. Moreover, any two partitions consisting of the same core of 3 species must have the same partition-productive cone, as the stoichiometric subspace is of dimension 3. For example, the partition-productive cones of $\{\{\bar{1}\}, \{\}, \{\bar{2}, \bar{3}, \bar{4}\}\}$ and $\{\{\}, \{\}, \{\bar{1}, \{\bar{2}, \bar{3}, \bar{4}\}\}\}$ are identical. The intersections of partition-productive cones for different FWMC partitions for the complete 1-constituent CCRN with $L = 4$ are shown in the Fig. 13.

Unlike the $L = 4$ case, for the complete 1-constituent CCRN with $L = 5$ of order two, there can be equivalent MASs with identical FWMC-partitions whose species-productive cones do not intersect. Moreover, it is not a priori clear whether the partition-productive cones of different FWMC partitions will intersect. This is why we employed our visualization scheme in Fig. 7. As described in Sec. IV C, for a particular equivalence class we also show the intersection information of the species-productive cones in Fig. 8.

2. $L=3$ 1-constituent CCRN

Consider the complete 1-constituent CCRN with $L = 4$ of order two, given by

$$\begin{aligned} \mathcal{H}_1^3 &= (\mathcal{S}, \mathcal{R}) \\ \mathcal{S} &= \{\bar{1}, \bar{2}, \bar{3}\} \\ \mathcal{R} &= \{\bar{1} + \bar{1} \xrightleftharpoons[k_2]{k_{11}} \bar{2}, \\ &\quad \bar{1} + \bar{2} \xrightleftharpoons[k_3]{k_{12}} \bar{3}, \\ &\quad \bar{1} + \bar{3} \xrightleftharpoons[k_{22}]{k_{13}} \bar{2} + \bar{2}\}. \end{aligned}$$

This network has two MASs, namely

$$\begin{aligned} A_1 &= \{\bar{1} + \bar{2} \rightarrow \bar{3}, \\ &\quad \bar{1} + \bar{3} \rightarrow \bar{2} + \bar{2}\}, \\ A_2 &= \{\bar{2} \rightarrow \bar{1} + \bar{1}, \\ &\quad \bar{3} + \bar{1} \rightarrow \bar{2} + \bar{2}\}. \end{aligned}$$

Since the CCRN has one conservation law, the stoichiometric subspace is of dimension 2 and perpendicular to $[1, 2, 3]^T$. The two-dimensional stoichiometric subspace is shown in Fig. 14. The 6 rays emanating from the origin, labelled by their directions in the 3–dimensional space, form the edges of productive cones where one species is consumed and another is produced. The species-productive cones of the two autocatalytic cycles are also labelled. Note that the species-productive regions is the

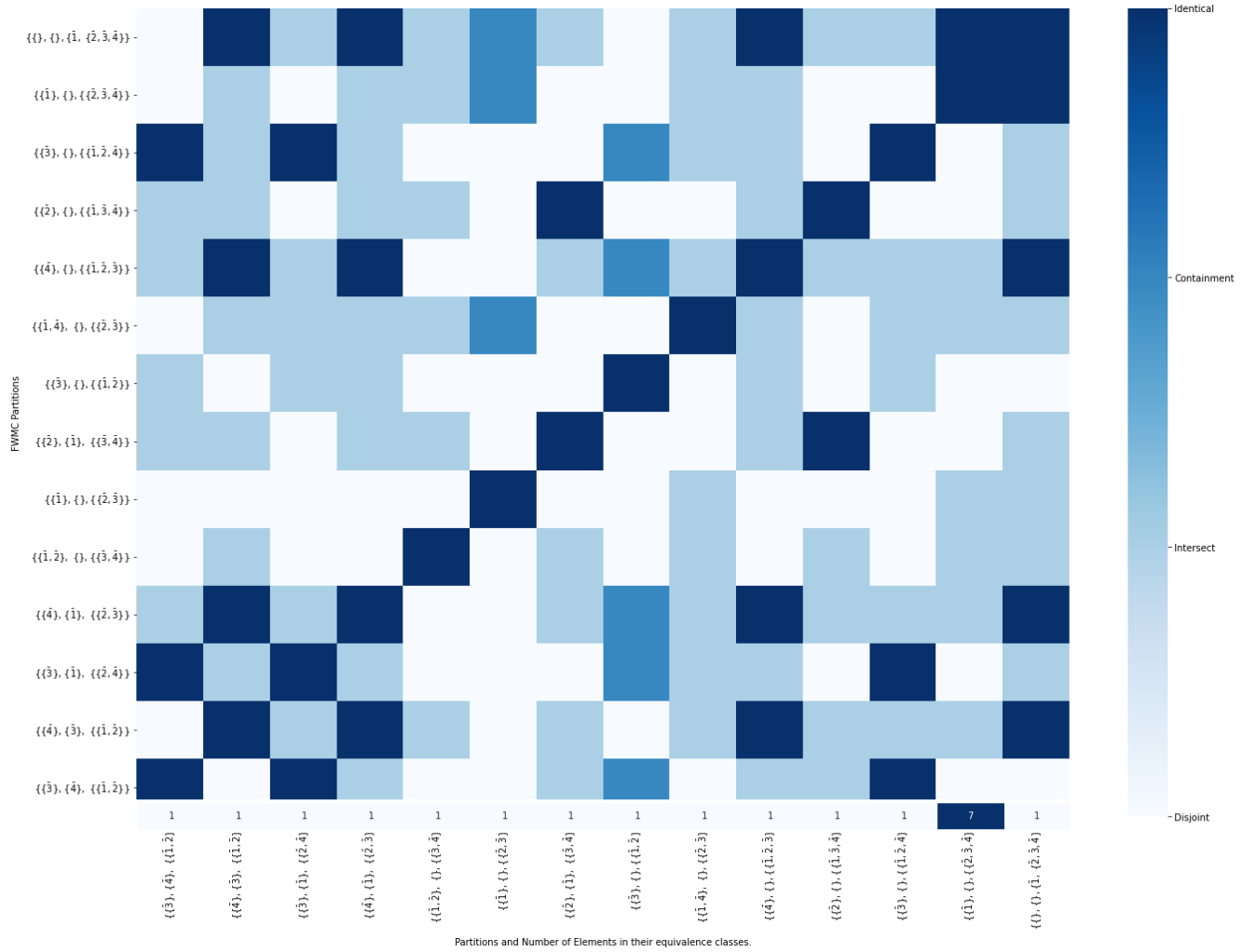


Figure 13: The list of the FWMC classes with the information of their intersection for the complete 1-constituent CCRN with $L = 4$ is shown.

complete cone within the bounding rays, and the finite boundary is drawn simply to enhance visualization. For example, the species-productive cone of the MAS A_1 is bounded by the rays $[-2, 1, 0]$ and $[-3, 0, 1]$. Notice that the FWMC partition of A_1 is $\{\{\bar{1}\}, \{\bar{1}\}, \{\bar{2}, \bar{3}\}\}$, consistent with the species-productive region that strictly consumes species $\bar{1}$ and strictly produces species $\bar{2}$ and $\bar{3}$.

We also want to make a few remarks about details that are not visible in Fig. 14. Firstly, for every MAS, there is a minimal drainable subnetwork obtained by reversing the reaction edges. The species-productive cones of these drainable networks will be the negatives of their autocatalytic counterparts. Secondly, notice that there is no MAS with the partition $\{\{\bar{2}\}, \{\bar{1}\}, \{\bar{1}, \bar{3}\}\}$. Had there been one, as would have been the case had we allowed reactions of order 3, then its species-productive cone would be bound by the rays $[0, -3/2, 1]$ and $[1, -1/2, 0]$.

Notice from Table I that the complete CCRN with $L = 3$ has a deficiency of one. This means that under mass-action kinetics, it has the potential to exhibit

multistability. To investigate this, we used Feinberg's deficiency one algorithm [30] in Appendix C and obtained rate constants for which the CCRN has multiple steady states. An example of multiple steady states and the rate constants using which they are obtained are reported in Fig. 15.

3. Number theoretic result for 1-constituent CCRN of order two

Since CCRNs are defined by number conservation laws, finding autocatalytic subnetworks in any CCRN corresponds to finding number-valued solutions of systems of equations. In this subsection we will show that if there exists a two-reaction MAS in a partition of a 1-constituent CCRN, then it must be unique.

Lemma V.1. *Let A be a 2-reactions MAS in a 1-constituent CCRN of order two. Then, there does not*

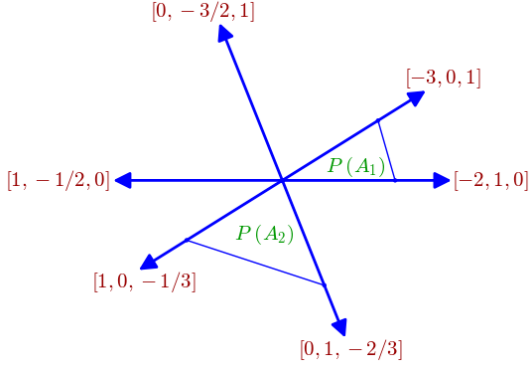


Figure 14: The 2-dimensional stoichiometric subspace along with the autocatalytic productive cones for the $L = 3$ complete 1-constituent CCRN are shown. The 6 rays emanating from the origin, labelled by their directions in the 3-dimensional space, form the edges of productive cones where one species is consumed and another is produced. For instance, the ray $[-3, 0, 1]$ corresponds to the direction in the stoichiometric space where $\bar{1}$ is consumed, $\bar{3}$ is produced, and $\bar{2}$ remains constant.

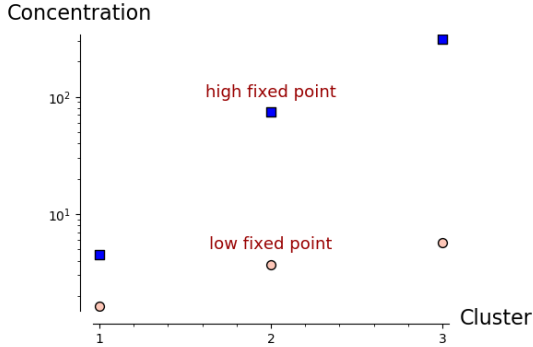


Figure 15: Multiple fixed points for the complete 1-constituent CCRN with $L = 3$ and order two obtained using Feinberg's deficiency one algorithm (see Appendix C). The parameters for the algorithm are chosen to be $\mu_{\bar{1}} = 1, \mu_{\bar{2}} = 3, \mu_{\bar{3}} = 4 = \eta$, and the rate constants for the model are $k_{11} = 1, k_2 = 1.001, k_{12} = 1.1, k_3 = 1.011, k_{13} = 0.146, k_{22} = 0.027$.

exist any other MASs in the CCRN with the same FWMC partition as A .

Proof. Let $A_C = \{\bar{i}, \bar{j}\}$ be the core species in A (assume without loss of generality that $i > j$). The possible 2-reaction MASs with that set of core species are given by

reactions of the form

$$\begin{cases} \bar{i} + \bar{\ell}_1 \rightarrow \bar{j} + \bar{\ell}_2, \\ \bar{j} + \bar{\ell}_3 \rightarrow 2\bar{i}. \end{cases}$$

$$\begin{cases} \bar{i} + \bar{m}_1 \rightarrow 2\bar{j}, \\ \bar{j} + \bar{m}_2 \rightarrow \bar{i} + \bar{m}_3. \end{cases}$$

Note that each non-core species can appear with the value 0, indicating the absence of a reactant. However, due to Lemma IV.3, the conservation law will require that at least one food species is non zero. Observe, also, that in both types of potential MASs there is a reaction which is uniquely determined by $\bar{i}$ and $\bar{j}$ (being then $\bar{\ell}_3$ and $\bar{m}_1$ also unique for this choice of the core set). Concretely $\bar{\ell}_3 = 2\bar{i} - \bar{j}$ and $\bar{m}_1 = 2\bar{j} - \bar{i}$. Note that these reactions are different since it is assumed that $i \neq j$. Thus, the above reactions reduce to:

$$M_1 : \begin{cases} \bar{i} + \bar{j} + \bar{\ell} - \bar{i} \rightarrow \bar{j} + \bar{\ell}, \\ \bar{j} + 2\bar{i} - \bar{j} \rightarrow 2\bar{i}. \end{cases}$$

$$M_2 : \begin{cases} \bar{i} + 2\bar{j} - \bar{i} \rightarrow 2\bar{j}, \\ \bar{j} + \bar{i} + \bar{m} - \bar{j} \rightarrow \bar{i} + \bar{m}. \end{cases}$$

Now we will do a case-by-case analysis to show that each FWMC-partition will contain at most one solution of the systems of equations.

Note that the only possible partition of the species induced by the reactions in M_1 is: $FWMC(M_1) = \{\bar{j} + \bar{\ell} - \bar{i}, 2\bar{i} - \bar{j}\}, \{\bar{\ell}\}, \{\{\bar{i}, \bar{j}\}\}$ whereas the partition for M_2 is: $FWMC(M_2) = \{2\bar{j} - \bar{i}, \bar{i} + \bar{m} - \bar{j}\}, \{\bar{m}\}, \{\{\bar{i}, \bar{j}\}\}$. From both types of partition, the only possibility for which both coincide is that $\bar{\ell} = \bar{m}$ and $\bar{i} = \bar{j}$ contradicting the assumption that $i > j$. $\square$

Given two different species $\bar{i}, \bar{j}$ with $i > j \geq 1$, the above result also provides a way to construct all the 2-reactions MASs with those species being the core species. The explicit reactions taking part of the MASs are in the form:

- In case $i \leq \frac{L}{2}$:

$$MAS_1 = \begin{cases} R_1 : \bar{i} + \bar{j} + \bar{\ell} - \bar{i} \rightarrow \bar{j} + \bar{\ell}, \\ R_2 : \bar{j} + 2\bar{i} - \bar{j} \rightarrow 2\bar{i}. \end{cases}$$

for all $\ell \in [i - j, L] \cap \mathbb{Z}$.

- In case $\frac{i}{2} \leq j \leq \frac{L}{2}$:

$$MAS_2 = \begin{cases} R_1 : \bar{i} + 2\bar{j} - \bar{i} \rightarrow 2\bar{j}, \\ R_2 : \bar{j} + \bar{i} + \bar{m} - \bar{j} \rightarrow \bar{i} + \bar{m}. \end{cases}$$

for all $m \in [0, L + j - i] \cap \mathbb{Z}$.

4. Computational challenges in scaling

Consider the complete 2-constituent CCRN of order two, defined as

$$\begin{aligned}\mathcal{H}_2 &= (\mathcal{S}, \mathcal{R}) \\ \mathcal{S} &= \{\overline{a, b} | a, b \in \mathbb{Z}_{\geq 0}\} \\ \mathcal{R} &= \{\overline{a, b} + \overline{c, d} \rightarrow \overline{e, f} + \overline{g, h} | \\ &\quad a + c = e + g \text{ and } b + d = f + h\}.\end{aligned}$$

We wish to remark that for $\mathcal{H}_2$ and other CRNs with a large number of reactions, generating the whole set of autocatalytic subnetworks is computationally challenging. As the number of reaction increases, the dimension of the space where the MASs are to be found also increases, and even finding a single MAS in the network implies solving a difficult optimization problem that might be computationally costly. Even if finding the MASs is polynomial-time solvable (which does not seem to be the case), the complexity of the enumeration algorithm may turn into exponential (for e.g., see [31]), since the number of subnetworks increases considerably with the number of species and reactions. Thus for practical reasons, it may be more feasible to consider sparser rule-generated networks, as explained in the next subsection and shown in Fig. II.

C. Rule generated CCRNs

A naive computational modelling of artificial chemistry [32] often runs into issues of memory, storage and processing due to a combinatorial explosion of chemical species and reactions involved in the CRN. For instance, if one uses string chemistry [33] (or polymer sequence chemistry) to model polymers of length up to N with D distinct monomers (D constituents), a straightforward calculation shows that one is required to track $\sum_{i=1}^N D^i > D^N$ species. Since for any physically relevant model $N \gg D$, so it is clear that the computation will soon become intractable as one increases the length of the string (polymer) N . Even a simple graph-grammar for generating artificial chemistry, for e.g. with the application of [34], can generate a reaction network that is computationally intractable. Moreover, even in the computationally tractable regime, an enumeration of autocatalytic subnetworks or motifs will generically yield an extremely long list. In such a scenario, it might be unclear how to simplify the model without stripping away the essential details or adding artifacts.

We argue that the *Cluster framework* can help alleviate some of the computational issues. Any rule generated CRN, in string chemistry or built using graph-grammar methods must induce rules on the CCRN. Since the induced CCRN can have an exponentially reduced species set, it offers a way to constrain the model to a more computationally tractable regime. Since autocatalytic

Monomer activation	$\bar{1} \rightleftharpoons \bar{1}^*$
Catalyzed activation	$\bar{C} + \bar{1} \rightleftharpoons \bar{1}^* + \bar{C}$
Polymerization	$\bar{1} + \bar{1}^* \rightleftharpoons \bar{2}$ $\bar{2} + \bar{1}^* \rightleftharpoons \bar{3}$ $\vdots$
1-constituent	$\bar{j} + \bar{1}^* \rightleftharpoons \overline{j+1}$
2-constituent	$\left\{ \begin{aligned} \overline{j, k} + \overline{1, 0}^* &\rightleftharpoons \overline{j+1, k} \\ \overline{j, k} + \overline{0, 1}^* &\rightleftharpoons \overline{j, k+1} \end{aligned} \right.$
$\vdots$	$\vdots$
Templated polymerization	$\left\{ \begin{aligned} \overline{j, k} + \overline{k-1, j+1, 0}^* &\rightleftharpoons \overline{j, k} + \overline{k, j} \\ \overline{j, k} + \overline{k, j-1+0, 1}^* &\rightleftharpoons \overline{j, k} + \overline{k, j} \end{aligned} \right.$
Fusion	$\bar{j} + \bar{k} \rightleftharpoons \overline{j+k}$
Fission	$\bar{j} \rightleftharpoons \bar{j-k} + \bar{k}$

Table II: Several examples of rules for generating a sparse CCRN relevant to biochemistry. Here, in accordance with the definition of a CCRN, the variables j, k must be chosen such that the resulting clusters have a non-negative integer quantity of each constituent.

networks are preserved under this coarsening (with some caveats mentioned in the introduction to this section), one can use the CCRN to obtain a more manageable list of MASs. One can then gradually complexify the CCRN by adding more species with the same conserved quantities until the desired behavior of the CRN is captured by the CCRN.

For example, consider polymer chemistry with two monomers A and B . For an arbitrary polymer ω , the addition of an A or B yields $\omega \cdot A$ or $\omega \cdot B$, respectively. In the cluster framework, if we map A, B , and ω to $\bar{1}, \bar{0}$, $\bar{0}, \bar{1}$, and $\bar{j}, \bar{k}$, respectively, the polymerization reactions in the CCRN become

$$\begin{aligned}\overline{j, k} + \overline{1, 0} &\rightleftharpoons \overline{j+1, k}, \\ \overline{j, k} + \overline{0, 1} &\rightleftharpoons \overline{j, k+1}.\end{aligned}$$

In realistic chemistry, however, a monomer addition to a polymer is only done by an activated monomer and not an unactivated monomer. Thus, we may want to distinguish activated and unactivated monomers in our model, for which we can add extra species $\bar{1}, \bar{0}^*$ and $\bar{0}, \bar{1}^*$. The resulting polymerization reactions and other examples of rule generated CCRNs are shown in Table II.

VI. DISCUSSION AND FUTURE RESEARCH

In this work, we began by presenting a linear-algebraic representation of CRNs in terms of an input-output matrix pair. We introduced different notions of autocatalysis existing in literature as nested classes of conditions on the CRN. In particular, the results and analysis in this work is for the class of *exclusive autocatalysis* [6], which is more encompassing of true cases than the notion of autocatalysis used by Blokhuis et al. in [11]. We defined minimal autocatalytic subnetworks (MASs), proved properties of their stoichiometric matrices, and explained their induced polyhedral geometry. Next, we proposed mathematical optimization based algorithms to exhaustively compute all the MASs in a CRN and create their visualizations. Finally, we introduced the notion of cluster CRNs (CCRN) as a useful coarse-graining induced by general CRNs, and employed our techniques on CCRNs obtained from one constituent. Notably, we show that the list of MASs for maximally connected 1-constituent CCRNs with length up to L increases exponentially with L .

The mathematical theory developed here facilitated the development of effective algorithms for identifying MASs in general CRNs. Nevertheless, our optimization-based approach requires solving a series of binary mathematical programming problems with an increasing number of linear constraints. This approach is computationally costly for large CRNs. However, there are reasons to hope that further research on these optimization problems, in particular, its polyhedral properties, will allow us to strengthen the formulations and solve them more efficiently. Exploiting the algebraic properties behind conservation laws for sequences of integers holds particular promise in this regard. We have also introduced a simple coarse graining of CRNs as CCRNs. By adjusting the number of dimensions included in a CCRN it is possible to trade-off computational tractability against fidelity to the complete CRN and to accurately capture all autocatalytic motifs. This flexibility is a great strength of the CCRN model. Additionally, although we here only considered clusters with nonnegative constituents, it also should be possible to consider clusters defined on the complete integer lattice. Such a reaction network could be used to model nuclear reactions where the positive conserved quantities correspond to positive charges or subatomic particles and the negative conserved quantities would refer to negative charges or antiparticles.

In our explorations we only considered fully connected CCRNs, but some reactions that are consistent with mass conservation are, nonetheless, impossible in real chemistry. In future work it would be interesting, similar to [35, 36], to consider randomly generated CCRNs in an Erdos-Renyi framework and compute a bound on the number of autocatalytic cycles that persist as a function of the probability of a reaction's being allowed in the network. An additional, and perhaps even more promising approach to representing real chemistry would be to ex-

pand our understanding of the kinetics of fully-connected CCRNs when rate constants differ over many orders of magnitude. Such an effort might be aided by understanding the effects of adding thermodynamic parameters such as free-energy assignments to different clusters.

Overall, the mathematical framework and computational tools developed here have potential applications for any kind of CRN that has multiple MASs: an autocatalytic ecosystem. This is of particular significance for better connecting biology and chemistry at multiple scales of analysis. As discussed by [37], a cell's ability to grow and reproduce rests on the fact that its metabolism is an autocatalytic ecosystem. Even ignoring autocatalytic motifs that include genetic mechanisms (i.e., those whose members are proteins or nucleic acids), metabolic networks contain many autocatalytic motifs [25]. The tools developed here will facilitate detecting MASs and their ecological interactions within metabolic networks, which has the potential to help elucidate diverse biochemical processes.

Studies of natural ecological communities, from coral reefs to rainforests, could also be informed by studies of autocatalytic CRNs. An organism (a cell or multicellular system) can be coarse-grained as a single autocatalytic motif composed of numerous lower-level cooperating MASs. Here, the chemical complex that corresponds to the organism is a member of an autocatalytic motif, which sits in a species-productive cone due to the input of food derived from the physical environment (e.g., light and carbon dioxide in the case of photosynthetic organisms) and/or other from other organisms. However, although it ought to be possible to apply the tools developed here to enrich community ecosystem, in practice the lack of stoichiometric data for ecology (e.g., how many rabbits react with one wolf to make two wolves?) make this challenging at the current time.

Even if we currently lack the data sets needed to mathematically analyze metabolisms or other biological autocatalytic ecosystems, the work here has important implications for studying origins of life. Biological autocatalysis is not discretely different from that found in non-biological CRNs. The continuity between the two implies that the origin of life can be viewed as a process of expansion and complexification of autocatalytic ecosystems via the progressive activation (i.e., transitioning to positive flux) of more and more MASs over time. In future work, we hope to introduce kinetics to CCRNs and expand the insights gained here to identify conditions that permit and promote the open-ended exploration of the attractor space associated with CCRNs in an autocatalytic ecosystem.

Acknowledgements

We want to thank Connor Simpson for help with combinatorics and programming, Alex Plum, Tymofii Sokolyskiy, and Zhen Peng for biological discussions, and Asvin

G., Dave Auckly, Diego Rojas, Gautam Neelakantham, Jeff Linderth, and Polly Yu for mathematical discussions. We additionally want to thank Vladimir Sotirov for a careful reading of the manuscript and numerous suggestions. Finally, we thank an anonymous reviewer for devoting time and effort to significantly improve the work.

Declarations

Competing interests The authors declare that they have no conflict of interest.

Authors' contributions Conceptualization: Praful Gagrani, Victor Blanco, Eric Smith, David Baum; Formal analysis and investigation: Praful Gagrani; Computational analysis and investigation: Victor Blanco; Writing - original draft preparation: Praful Gagrani, Victor Blanco; Writing - review and editing: Praful Gagrani, Victor Blanco, David Baum, Eric Smith; Funding acquisition: David Baum, Eric Smith, Victor Blanco.

Funding This work was partially funded by the National Science Foundation, Division of Environmental Biology (Grant No: DEB-2218817). V. Blanco acknowledges the financial support by the Spanish Ministerio de Ciencia e Innovación AEI/FEDER grant number PID2020-114594GBC21 and IMAG-Maria de Maeztu grant CEX2020-001105-M /AEI /10.13039/501100011033, Junta de Andalucía projects P18-FR-1422/2369, AT 21_00032, FEDERUS-1256951, B-FQM-322-UGR20, and UE-NextGenerationEU (ayudas de movilidad para la recualificación del profesorado universitario).

Availability of data and materials The Python codes for detecting minimal autocatalytic subnetworks are made available at <https://github.com/vblanco0R/autocatatalyticsubnetworks> [38].

Appendix A: Deficiency theory and multistability in CRNs

In [13], a CRN is defined by the triple $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$, where $\mathcal{S}$, $\mathcal{C}$, and $\mathcal{R}$ are the sets of species, complexes, and reactions respectively. Recall from Sec. IIB that a CRN is a hypergraph with hypervertices in the set $\mathcal{C}$ and hyperedges in $\mathcal{R}$. The stoichiometric matrix $\mathbb{S}$ is defined such that (see Eq. 4)

$$\text{cols}(\mathbb{S}) = \{y' - y | y \rightarrow y' \in \mathcal{R}\}.$$

Alternatively, $\mathbb{S}$ can be written as the product $Y \cdot \mathbb{M}$, where the columns of Y are the stoichiometries of the hypervertices

$$\text{cols}(Y) = \{y | y \in \mathcal{C}\}$$

and $\mathbb{M}$ is the oriented vertex-edge incidence matrix of the hypergraph (see [13, 39])

$$\text{cols}(\mathbb{M}) = \{e_{y'} - e_y | y \rightarrow y' \in \mathcal{R}\},$$

where e_y is the vector with 1 at hypervertex y and 0 otherwise. Note that Y and $\mathbb{M}$ are of sizes $|\mathcal{S}| \times |\mathcal{C}|$ and $|\mathcal{C}| \times |\mathcal{R}|$, respectively, yielding $\mathbb{S}$ of size $|\mathcal{S}| \times |\mathcal{R}|$.

Since $\mathbb{S}$ is a linear operator, using the rank-nullity theorem we get

$$\dim(\text{domain}(\mathbb{S})) = \dim(\text{im}(\mathbb{S})) + \dim(\text{ker}(\mathbb{S})). \quad (\text{A1})$$

Let E, V, s denote the number of reactions (hyperedges), complexes (hypervertices) and the dimension of the stoichiometric subspace S , respectively. Then, by definition,

$$\begin{aligned} \dim(\text{domain}(\mathbb{S})) &= E, \\ \dim(\text{im}(\mathbb{S})) &= s. \end{aligned}$$

Using $\mathbb{S} = Y \cdot \mathbb{M}$, we have

$$\dim(\text{ker}(\mathbb{S})) = \dim(\text{ker}(\mathbb{M})) + \dim(\text{ker}(Y) \cap \text{im}(\mathbb{M})). \quad (\text{A2})$$

Notice that the complexes and reactions, without referring to the underlying species, form a graph, the vertex-edge incidence matrix of which is given by $\mathbb{M}$. Using properties of the vertex-edge incidence matrix, if the number of independent loops in the graph are ι , then

$$\dim(\text{ker}(\mathbb{M})) = \iota.$$

Denoting $\dim(\text{ker}(Y) \cap \text{im}(\mathbb{M}))$ by δ , also referred to as the *deficiency* of the CRN, substituting the above and Eq. A2 in Eq. A1, we get

$$\delta = E - \iota - s. \quad (\text{A3})$$

Furthermore, let us denote the number of connected components or *linkage classes* of the complex-reaction graph by ℓ . Notice that ℓ and ι are the zeroth and first Betti numbers of a graph, respectively [40]. Using the formula for the Euler characteristic of a graph, we have the relation

$$V - E = \ell - \iota, \quad (\text{A4})$$

where V is the number of vertices in the graph. Substituting the above in Eq. A3, we obtain the more widely used expression for the deficiency

$$\delta = V - \ell - s. \quad (\text{A5})$$

For direct derivations of Eq. A5 without resorting to a counting of independent loops, see [13, 18].

Intuitively, deficiency is a topological property of the reaction network that informs about the steady state dynamics. In particular, a non-zero deficiency is a necessary, but not sufficient, condition for the dynamics to

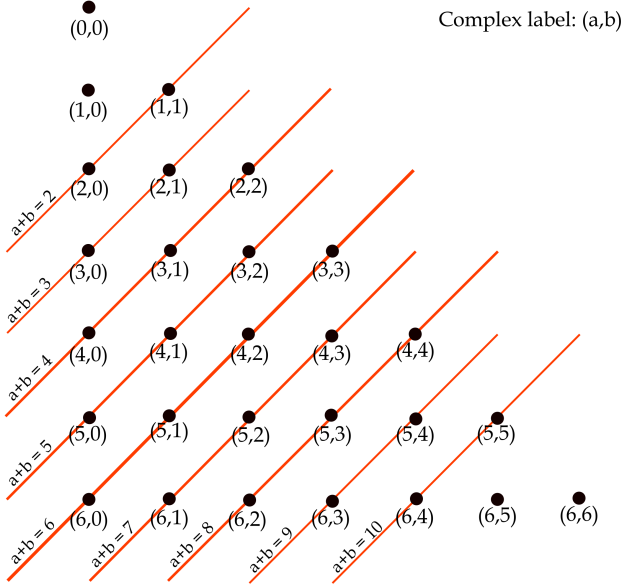


Figure 16: Representation of the complete 1-constituent CCRN with $L = 6$. Complex label (a, b) on a lattice point corresponds to the complex $\bar{a} + \bar{b}$. All the complexes connected by a red line correspond to a fully connected graph in the reaction network. For example, the red line connecting $\{(5, 0), (4, 1), (3, 2)\}$, labelled $a + b = 5$, corresponds to the reactions $\{\bar{5} \rightleftharpoons \bar{4} + \bar{1} \rightleftharpoons \bar{3} + \bar{2} \rightleftharpoons \bar{5}\}$ in the CCRN.

exhibit *multistability* or multiple steady states. In certain cases, such as for deficiency one networks [30], there exist algorithms that can help classify whether or not multiple steady states can exist for a given CRN and identify the rate constants such that under mass-action kinetics it exhibits multistability. Once a model exhibiting multistability is finalized, one can also use algorithms from [19, 41] to estimate the most-likely path the system will take, along with its probability, to escape from one steady state to another.

Appendix B: Kernel of the stoichiometric matrix for the complete 1-constituent CCRN

Let $\mathcal{H}_1^L = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ denote a complete 1-constituent CCRN of length L and order two. 1-constituent CCRN of length L means that the species set consists of clusters up to length L , denoted as $\mathcal{S} = \{\bar{1}, \bar{2}, \dots, \bar{L}\}$ (for a detailed definition, see Sec. VB 1.) For a diagrammatic representation of the reaction network, see Fig. 16. Here, following Appendix A, we will calculate the kernel of the stoichiometric matrix for a general length L .

First, we will count the number of complexes in a complete 1-constituent CCRN of length L . Let us denote the set of complexes of size N as $\mathcal{C}_L^N$. Recall that the size

per constituent of a complex is defined as the sum of the conserved quantities in each constituent. For $2 \leq N \leq L$,

$$\mathcal{C}_L^N = \{\bar{N}, \bar{N-1} + \bar{1}, \dots, \bar{\lfloor N/2 \rfloor} + \bar{\lceil N/2 \rceil}\}.$$

For $L < N < 2L - 2$,

$$\mathcal{C}_L^N = \{\bar{N-L} + \bar{L}, \dots, \bar{\lfloor N/2 \rfloor} + \bar{\lceil N/2 \rceil}\}.$$

Thus, the number of complexes as a function of N and L are

$$|\mathcal{C}_L^N| = \begin{cases} \lfloor N/2 \rfloor + 1 & \text{for } 2 \leq N \leq L, \\ L + 1 - \lceil N/2 \rceil & \text{for } L < N \leq 2L - 2. \end{cases}$$

The total number of complexes for a CCRN of length L is thus

$$|\mathcal{C}_L| = \sum_N |\mathcal{C}_L^N| = \frac{(L+1)(L+2)}{2} - 4.$$

This can be found by either summing the individual contributions or by observing from Fig. 16 that the complexes occupy the lower triangle including the diagonal for a lattice with $L + 1$ rows and columns excluding 4 points.

Let us denote the set of reactions of complexes of size N by $\mathcal{R}_L^N$. Since the reaction network of the complete CCRN is obtained by connecting all allowed reactions, for complexes of size N , the number of reactions is given by

$$|\mathcal{R}_L^N| = \begin{cases} 2 \binom{\lfloor N/2 \rfloor + 1}{2} & \text{for } 2 \leq N \leq L, \\ 2 \binom{L+1 - \lceil N/2 \rceil}{2} & \text{for } L < N \leq 2L - 2, \end{cases}$$

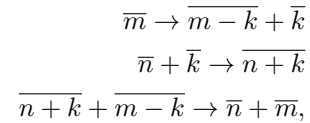
where the factor of 2 is multiplied since all reactions are reversible. Since, each N contributes exactly one linkage class (or connected component), the total number of linkage classes ℓ for the complete CCRN is $2L - 3$.

Recall from Appendix A, the kernel of the stoichiometric matrix is the sum of the independent loops ι and the deficiency δ , where using Equations A4 and A5 we have

$$\iota = |\mathcal{R}_L| - |\mathcal{C}_L| + \ell \quad (\text{B1})$$

$$\begin{aligned} &= \sum_{N=2}^{2L-2} |\mathcal{R}_L^N| - \frac{L^2 - 5L - 2}{4} \\ \delta &= |\mathcal{C}_L| - \ell - s \quad (\text{B2}) \\ &= \frac{(L+1)(L+2)}{2} - 4 - (2L-3) - (L-1) \\ &= \frac{(L-1)(L-2)}{2}. \end{aligned}$$

The intersection $\text{Im}(\mathbb{M}) \cap \text{Ker}(Y)$, the dimension of which is the deficiency δ , counts the null flows of the CCRN which are not simple loops in the network. While an explicit basis can be obtained using subnetworks of the form

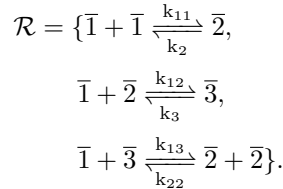


we will not be pursuing it since we do not need it for any result in this work.

Appendix C: Deficiency-one algorithm for $L=3$ 1-constituent CCRN

The complete $L = 3$ 1-constituent CCRN is the simplest complete CCRN, and as we will see, it possesses non-trivial dynamical properties. In particular, from Table I, it can be seen that the deficiency δ for the CCRN is one. This hints at the possibility that the CRN, when taken with mass-action kinetics, exhibits multistability. In this section, using Feinberg's deficiency-one algorithm from [13, 30], we will confirm that the network can indeed exhibit multiple steady states, and find a relation between the rate constants and concentrations where that is the case. The algorithm works by converting the CRN into a system of linear inequalities, and if they have a solution, they can be used to obtain a relation between the rate constants and multiple steady states.

The $L = 3$ 1-constituent CCRN, denoted by $\mathcal{G} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$, is given by the species set $\mathcal{S} = \{\bar{1}, \bar{2}, \bar{3}\}$, complex set $\mathcal{C} = \{\bar{1} + \bar{1}, \bar{2}, \bar{1} + \bar{2}, \bar{3}, \bar{1} + \bar{3}, \bar{2} + \bar{2}\}$ and reaction set



Notice that the above reaction network is a regular [30] deficiency-one reaction network, and is thus a valid candidate for the deficiency-one algorithm. To proceed with our calculation, we will follow closely the example in Chapter 17 of [13] and Sec. 5 in [30].

The confluence vector $g \in \mathbb{R}^{\mathcal{C}}$ for our network can be calculated to be (up to sign)

$$\begin{aligned}g_{\bar{2}} &= -1, & g_{\bar{1}+\bar{1}} &= 1, \\ g_{\bar{1}+\bar{2}} &= -1, & g_{\bar{3}} &= 1, \\ g_{\bar{1}+\bar{3}} &= -1, & g_{\bar{2}+\bar{2}} &= 1.\end{aligned}$$

For our construction, we will use the following upper-middle-lower partition

$$\begin{aligned}U &:= \{\bar{1} + \bar{3}, \bar{2} + \bar{2}\}, \\ M &:= \{\bar{1} + \bar{2}, \bar{3}\}, \\ L &:= \{\bar{1} + \bar{1}, \bar{2}\}.\end{aligned}$$

The partition induces the following system of inequalities on the chemical potential,

$$2\mu_{\bar{2}} > \mu_{\bar{1}} + \mu_{\bar{3}} > \mu_{\bar{3}} = \mu_{\bar{1}} + \mu_{\bar{2}} > \mu_{\bar{2}} > 2\mu_{\bar{1}}.$$

The inequalities are satisfied by

$$\begin{aligned}\mu_{\bar{1}} &> 0, \\ \mu_{\bar{2}} &> 2\mu_{\bar{1}}, \\ \mu_{\bar{3}} &= \mu_{\bar{1}} + \mu_{\bar{2}},\end{aligned}$$

which clearly have a solution, for e.g. $\mu_{\bar{1}} = 1, \mu_{\bar{2}} = 3, \mu_{\bar{3}} = 4$. Thus, the CCRN can exhibit multiple steady states.

To find the rate constants at which the system exhibits multistability, we use subsection 5.3 in [30]. Following the reference, let us denote the mass-action kinetics rate vector by $\kappa_{y \rightarrow y'} := k_{y \rightarrow y'} x^y$, where x is the concentration vector and $x^y := \prod_{i \in \mathcal{S}} x_i^{y_i}$. Choosing the monotonically increasing function to be the exponential $\phi(x) = e^x$, and choosing η so that $\mu_{\bar{1}} + \mu_{\bar{3}} < \eta < \mu_{\bar{2}}$ we get

$$\begin{aligned}\kappa_{\bar{1}+\bar{3} \rightarrow \bar{2}+\bar{2}} &= 1 \cdot \frac{e^{2\mu_{\bar{2}}} - e^{\eta}}{e^{2\mu_{\bar{2}}} - e^{\mu_{\bar{1}}}\mu_{\bar{3}}} = k_{13}x_{\bar{1}}x_{\bar{3}}, \\ \kappa_{\bar{2}+\bar{2} \rightarrow \bar{1}+\bar{3}} &= 1 \cdot \frac{e^{\mu_{\bar{1}}+\bar{3}} - e^{\eta}}{e^{2\mu_{\bar{2}}} - e^{\mu_{\bar{1}}}\mu_{\bar{3}}} = k_{22}x_{\bar{2}}^2, \\ \kappa_{\bar{1}+\bar{2} \rightarrow \bar{3}} &= k_{12}x_{\bar{1}}x_{\bar{2}}, \\ \kappa_{\bar{3} \rightarrow \bar{1}+\bar{2}} &= k_{12}x_{\bar{1}}x_{\bar{2}} - 1 = k_3x_{\bar{3}}, \\ \kappa_{\bar{2} \rightarrow \bar{1}+\bar{1}} &= 1 \cdot \frac{e^{2\mu_{\bar{1}}} - e^{\eta}}{e^{2\mu_{\bar{1}}} - e^{\mu_{\bar{2}}}} = k_2x_{\bar{2}}, \\ \kappa_{\bar{1}+\bar{1} \rightarrow \bar{2}} &= 1 \cdot \frac{e^{\mu_{\bar{2}}} - e^{\eta}}{e^{2\mu_{\bar{1}}} - e^{\mu_{\bar{2}}}} = k_{11}x_{\bar{1}}^2.\end{aligned}$$

From [30], it is also known that if $x := (x_{\bar{1}}, x_{\bar{2}}, x_{\bar{3}})$, then so is $x^* := (x_{\bar{1}}e^{\mu_{\bar{1}}}, x_{\bar{2}}e^{\mu_{\bar{2}}}, x_{\bar{3}}e^{\mu_{\bar{3}}})$, thus yielding multiple steady states for our system.

Appendix D: Details of the algorithm for detecting stoichiometrically autocatalytic motifs

In Section IV B, we give the main ideas behind our algorithm to construct exclusively autocatalytic subnetworks using the input and output stoichiometric matrices. In what follows we detail the algorithm that we propose to enumerate the entire list of minimal *stoichiometrically* autocatalytic subnetworks of a CRN, and that we applied to the CCRNs in our computational experiments.

Notation D.1. In this section, we abbreviate *minimal stoichiometrically autocatalytic subnetwork* as MAS.

The goal of the algorithm that we propose is to determine sets of core species and reactions conforming to the requirements of a MAS by solving, sequentially, a series of nested mathematical optimization problems. Apart from the variables y, z and $\mathbf{v}$ already described in Section IV B, we use the following set of variables in our Mixed-Integer Linear Programming (MILP) problem:

$$q_i = \begin{cases} 1 & \text{if species } i \text{ is a non-core present species} \\ & \text{in the MAS,} \\ 0 & \text{otherwise} \end{cases},$$

and Δ_i is a constant to be defined later for all $i \in \mathcal{S}$.

The MASTER optimization problem to construct MASs in this case is the following:

$$\min \sum_{r \in \mathcal{R}} z_r \quad (D1)$$

$$v_r \leq z_r, \forall r \in \mathcal{R}, \quad (D2)$$

$$\sum_{r \in \mathcal{R}} z_r \geq 2, \quad (D3)$$

$$y_i + q_i \leq 1, \forall i \in \mathcal{S}, \quad (D4)$$

$$y_i + q_i \geq z_r, \forall r \in \mathcal{R}, i \in \mathcal{S} \text{ with } S_i^r > 0 \neq 0, \quad (D5)$$

$$\sum_{\substack{i \in \mathcal{S}: \\ S_i^r > 0}} y_i \geq z_r, \forall r \in \mathcal{R}, \quad (D6)$$

$$\sum_{\substack{i \in \mathcal{S}: \\ S_i^r < 0}} y_i \geq z_r, \forall r \in \mathcal{R}, \quad (D7)$$

$$\sum_{\substack{r \in \mathcal{R}: \\ |S_i^r| > 0}} z_r \geq q_i, \forall i \in \mathcal{S}, \quad (D8)$$

$$\sum_{\substack{r \in \mathcal{R}: \\ S_i^r > 0}} z_r \geq y_i, \forall i \in \mathcal{S}, \quad (D9)$$

$$\sum_{\substack{r \in \mathcal{R}: \\ S_i^r < 0}} z_r \geq y_i, \forall i \in \mathcal{S}, \quad (D10)$$

$$\sum_{r \in \mathcal{R}} S_i^r v_r \geq \varepsilon - \Delta_i(1 - y_i), \forall i \in \mathcal{S}, \quad (D11)$$

$$y_i, q_i \in \{0, 1\}, \forall i \in \mathcal{S}, \quad (D12)$$

$$v_r \in [0, 1], z_r \in \{0, 1\} \forall r \in \mathcal{R}. \quad (D13)$$

There, the objective function accounts for minimizing the number of reactions in the MAS. This objective assures, that each of the generated autocatalytic subnetworks is support-inclusion minimal. Constraints (D2) allow the correct representation of variable z_r (if $v_r > 0$, then z_r is forced to take value one, if $v_r = 0$, by the minimization criterion, z_r will also take value 0). Constraint (D3) avoids solutions with a single reaction. Constraints (D4) assure that species are either core species or present non-core species but not both (it is also possible that they do not participate in the MAS). In case $y_i + q_i = 1$, species i is involved in the MASs either as food, member, core or waste species. Otherwise, species i is not a part of the MAS. Constraints (D5) assure that in case a reaction r is selected to be part of the MAS ($z_r = 1$), the reactants and product species of reaction r must be present (either as core species or present non-core species). Since q_i and y_i are binary variables, the sum of them being greater than one (but because of the previous constraint

the equality holds) implies that either i is a core species ($y_i = 1$), or non-core present species ($q_i = 1$). Otherwise, the constraint is redundant. Constraints (D6) and (D7) ensure that in case a reaction is in the MAS, there must be a core species as a reactant and another as a product in the restricted stoichiometric matrix. Observe that in case $z_r = 1$, the sum of the binary variables y_i with i being reactant/product of r must be at least one, implying that at least one of the reactant/product species must be activated as core species. In case $z_r = 0$, the constraint is redundant. Similarly, Constraints (D8) enforces that in case a species is selected to be a present non-core species ($q_i = 1$), then there must be a reaction where it is present. (D11) is the productivity condition (see Remark 6). There, in case $y_i = 1$ (i is a core species in the motif), the right-hand side in the constraint becomes 1, implying that $\sum_{r \in \mathcal{R}} S_i^r v_r \geq \varepsilon > 0$ (here ε is a given positive but small enough constant to assure that the production is positive).

Constants Δ_i , for $i \in \mathcal{S}$, are assumed to be big enough assuring that in case i is a non-core (present or not) species ($y_i = 0$), the *production* of species i with flow vector $\mathbf{v}$ must always be greater than $\varepsilon - \Delta_i$ (making the constraint redundant). Specifically, for those non-core species whose production is negative (as food species), the absolute value production is upper bounded by Δ_i . We take $\Delta_i = -\sum_{\substack{r \in \mathcal{R}: \\ S_i^r < 0}} S_i^r$, for all $i \in \mathcal{S}$.

The above model belongs to the class of discrete optimization problems, which are known to be computationally challenging [42], but at the same time tremendously useful in determining solutions of complex decision problems in various fields [43]. Although the problem of detecting a MAS is known to be a NP-complete [44], there are several available softwares with implemented routines to obtain solutions of integer programming problems (such as Gurobi, CPLEX, ...) in reasonable CPU time for medium size instances.

In order to speed up the solution approach, the above master problem can be strengthened by adding valid inequalities that reduce the search of solutions. For instance, as shown in Theorem III.4, since the number of reactions in an MAS coincides with the number of core species that are part of it, one can add this (redundant) constraint to the model:

$$\sum_{i \in \mathcal{S}} y_i = \sum_{i \in \mathcal{R}} z_j$$

Furthermore, in a reversible CRN, denoting by

$$\text{Rev} = \{(r, r') \in \mathcal{R} \times \mathcal{R} : \text{reaction } r \text{ is the reverse of reaction } r'\},$$

one must also impose the following constraints that avoid using a reaction and its reverse in the same subnetwork:

$$z_r + z_{r'} \leq 1, \forall (r, r') \in \text{Rev}.$$

That is, if r and r' are reverse reaction of each other, the z -variables of those reactions cannot be both one.

The solution of the problem above is a MAS with minimum number of reactions. Once the problem is solved, with optimal values in the variables $(\bar{y}, \bar{q}, \bar{z}, \bar{\mathbf{v}})$, the set of core species is obtained as:

$$\text{Core} = \{i \in \mathcal{S} : \bar{y}_i = 1\},$$

and the set of reactions involved in the MAS is:

$$\mathcal{R}' = \{r \in \mathcal{R} : \bar{z}_r = 1\}.$$

Among the set of present non-core species, $\mathcal{Q} = \{i \in \mathcal{N} : \bar{q}_i = 1\}$, one can classify them based on the sign of the entries of the restricted stoichiometric matrix:

$$\begin{aligned} \text{Food} &= \{i \in \mathcal{Q} : \mathcal{S}_i^r \leq 0, \forall r \in \mathcal{R}'\}, \\ \text{Waste} &= \{i \in \mathcal{Q} : \mathcal{S}_i^r \geq 0, \forall r \in \mathcal{R}'\}, \\ \text{Member} &= (\mathcal{Q} \cup \text{Core}) \setminus (\text{Food} \cup \text{Waste}) \end{aligned}$$

The algorithm to generate the whole set of MASs is described in Algorithm 1. There, at each iteration, ℓ , the above Master problem is solved. Once a solution is obtained, say $(y^\ell, q^\ell, z^\ell, \mathbf{v}^\ell)$, the tuple of reactions that are part of the MAS is considered to restrict the next MASs. In particular, it is disallowed for the next MAS to use the same reactions tuples, which assures the minimality of the generated MASs. This restriction is incorporated in the model by means of Constraint (D14). The process is repeated until no more autocatalytic subnetwork can be found respecting the pool of incompatibilities that have been added during the procedure. With this algorithm the MASs are generated while being sorted by number of reactions involved in the subnetwork since each time the problem is solved the number of reactions is minimized.

Data: $\mathcal{S}, \mathcal{R}, \mathbb{S}, \ell = 0$
while FEASIBLE **do**
 Solve MASTER: $\text{MAS}^\ell := (y^\ell, q^\ell, z^\ell, \mathbf{v}^\ell)$.
 if MASTER is feasible **then**
 Save the solution and add to MASTER the constraint:

$$\sum_{\substack{r \in \mathcal{R}: \\ z_r = 1}} z_j^\ell \leq |\{r \in \mathcal{R} : z_r^\ell = 1\}| - 1 \quad (\text{D14})$$

 $\ell + 1 \leftarrow \ell$.
 end
end
Result: Set of MASs of a CRN: $\text{MAS}^1, \dots, \text{MAS}^\ell$.

Algorithm 1: Optimization-based approach to enumerate MASs.

One of the advantages of using mathematical optimization-based approaches to determine *special* sets from an underlying set (as MASs from a CRN) is its flexibility to be adapted to different features. For instance, one can easily filter the obtained MASs to have

at most a number of reactions, MAXREACTIONS, involved by adding the following constraint to the model:

$$\sum_{r \in \mathcal{R}} z_r \leq \text{MAXREACTIONS}.$$

Analogously, one can also generate MASs where just a given set of species takes part in the core species set, $\mathcal{S}_0 \subseteq \mathcal{S}$, by adding:

$$\sum_{i \in \mathcal{S}_0} y_i = 0.$$

As already mentioned, computing the entire list of MASs in a CRN is computationally challenging. Each iteration in the proposed procedure requires solving a mixed integer linear optimization problem with the constraints in the initial Master problem, but also those in the shape of (D14) that are added each time a MASs is found. Thus, the difficulty of solving this optimization problem increases with the number of MASs for a given CRN. In Figure (17) we show the performance of the cumulative CPU time required to generate the MASs for 1-constituent CCRNS with $L = 4, 5$, and 6. As can be observed from the plots, the consumed CPU time increases linearly in the number of MASs, with a slope strictly greater than one.

Reaction id	Reaction
R1/R2	$2\bar{1} \rightleftharpoons \bar{2}$
R3/R4	$\bar{1} + \bar{2} \rightleftharpoons \bar{3}$
R5/R6	$\bar{1} + \bar{3} \rightleftharpoons \bar{4}$
R7/R8	$\bar{1} + \bar{4} \rightleftharpoons \bar{2} + \bar{3}$
R9/R10	$\bar{1} + \bar{3} \rightleftharpoons 2\bar{2}$
R11/R12	$2\bar{2} \rightleftharpoons \bar{4}$
R13/R14	$\bar{2} + \bar{4} \rightleftharpoons 2\bar{3}$

Table III: List of reactions for a CCRN with 1 constituent and $L = 4$.

1. Checking intersection of cones

Once the whole set of MASs has been generated, $\mathcal{M} = \{\text{MAS}_1, \dots, \text{MAS}_K\}$, we have also developed an optimization-based methodology to check the pairwise intersection of the different polyhedral flow-productive cones and partition-productive cones induced from the set $\mathcal{M}$. Consider two of these cones in $\mathbb{R}_+^d$, say $C_1(v_1, \dots, v_s)$ and $C_2(w_1, \dots, w_r)$, described by means of its generating vectors (see Section III C). We first check if they the interior of the cones intersect or not by finding a non negative vector that belongs to the intersection of the closed cones maximizing the sum of the entries of that vector with respect to each of the generators of the

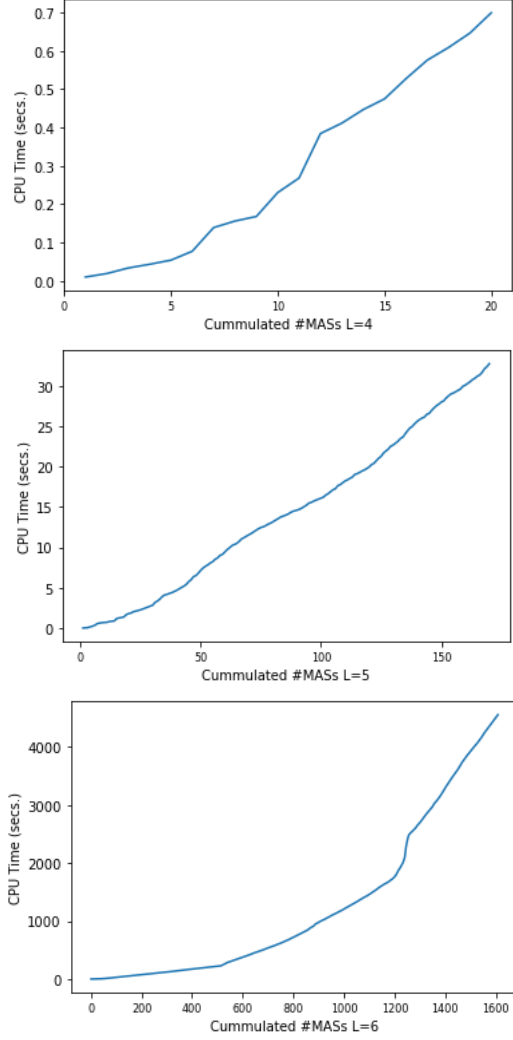


Figure 17: CPU time vs. # MASs for 1-constituent CCRN.

cones, i.e.:

$$\begin{aligned}
 \rho^*(C_1, C_2) &:= \max \sum_{l=1}^s \lambda_l + \sum_{l=1}^r \mu_l \\
 \text{s.t. } x &= \sum_{l=1}^s \lambda_l v_l, \\
 x &= \sum_{l=1}^r \mu_l w_l, \\
 x &\in \mathbb{R}_+^d, \\
 \lambda_l &\in [0, 1], \forall l = 1, \dots, s, \\
 \mu_l &\in [0, 1], \forall l = 1, \dots, r.
 \end{aligned}$$

If the optimal solution of the above linear programming problem is $\rho^*(C_1, C_2) = 0$, then, the cones do not inter-

sect (otherwise there should be another vector with greatest coefficients in the representation). If $\rho^*(C_1, C_2) > 0$, then, the cones intersect.

In case the cones intersect, different situations are possible: the cones are identical, one of the cones is strictly contained in the other, or the cones partially intersect. To detect each of the above possibilities, we compute the distances between each of the normalized generating vectors of one of the cones to the generating vectors in the other cones:

$$\phi_l^1 = D(v_l, \{w_1, \dots, w_r\}) = \min_{j=1, \dots, r} \{\|v_l - w_j\|\}$$

for $l = 1, \dots, s$, and

$$\phi_l^2 = D(w_l, \{v_1, \dots, v_s\}) = \min_{j=1, \dots, s} \{\|w_l - v_j\|\}$$

for $l = 1, \dots, r$.

With these values, in case the cones intersect:

- If $\sum_{l=1}^s \phi_l^1 = \sum_{l=1}^r \phi_l^2 = 0$, then for each of the generating vectors in both cones, there is a normalized generating vector of the other cone at distance zero from it. Thus, both cones are identical.
- If $\sum_{l=1}^s \phi_l^1 = 0$, but $\sum_{l=1}^r \phi_l^2 > 0$, then, $C_1 \subset C_2$ since for all the generating vectors in C_1 , there exist a generating vector in C_2 at distance zero from it.
- If $\sum_{l=1}^s \phi_l^1 > 0$, but $\sum_{l=1}^r \phi_l^2 = 0$, then, $C_2 \subset C_1$ by the same reasoning as above.
- If $\sum_{l=1}^s \phi_l^1 > 0$ and $\sum_{l=1}^r \phi_l^2 > 0$, then, the cones intersect but only partially.

Appendix E: Minimal Autocatalytic Subnetworks for CCRN with $L = 4$ and $L = 5$

Tables III and IV show the set of feasible reactions for 1-constituent CCRNs with $L = 4$ and $L = 5$, respectively. For those CCRNs we report in Tables V, VI, and VII the list of MASs obtained with our approach. Each row in those tables represent a MAS. In the first column of those tables (id) indicates the identifier of the MAS. The second column (Reactions (Flow)) provides the list of reactions taking part of the MAS. In parenthesis we provide a feasible flow for autocatalysis. In the third column (FWMC Partition), we report the FWMC partition for such a MAS. Finally, in the fourth column (Production) we show the production vector for the feasible flow in the second column, for each of the core species in the MAS.

Reaction id	Reaction
R1/R2	$2\bar{1} \rightleftharpoons \bar{2}$
R3/R4	$\bar{1} + \bar{2} \rightleftharpoons \bar{3}$
R5/R6	$\bar{1} + \bar{3} \rightleftharpoons \bar{4}$
R7/R8	$\bar{1} + \bar{4} \rightleftharpoons \bar{5}$
R9/R10	$\bar{1} + \bar{5} \rightleftharpoons 2\bar{3}$
R11/R12	$\bar{1} + \bar{5} \rightleftharpoons \bar{2} + \bar{4}$
R13/R14	$\bar{1} + \bar{4} \rightleftharpoons \bar{2} + \bar{3}$
R15/R16	$\bar{1} + \bar{3} \rightleftharpoons 2\bar{2}$
R17/R18	$2\bar{2} \rightleftharpoons \bar{4}$
R19/R20	$\bar{2} + \bar{3} \rightleftharpoons \bar{5}$
R21/R22	$\bar{2} + \bar{5} \rightleftharpoons \bar{3} + \bar{4}$
R23/R24	$\bar{2} + \bar{4} \rightleftharpoons 2\bar{3}$
R25/R26	$\bar{3} + \bar{5} \rightleftharpoons 2\bar{4}$

Table IV: List of reactions for a CCRN with 1 constituent and $L = 5$.

MAS id	Reactions (Flow)	FWMC Partition	Production
1	{R2(1), R9(1)}	{{3}, {}, {{1, 2}}}	(1, 1)
2	{R9(1), R13(1)}	{{1, 4}, {}, {{2, 3}}}	(1, 1)
3	{R4(3), R13(2)}	{{4}, {1}, {{2, 3}}}	(1, 1)
4	{R3(3), R9(2)}	{{1}, {}, {{2, 3}}}	(1, 1)
5	{R8(3), R13(2)}	{{2}, {1}, {{3, 4}}}	(1, 1)
6	{R5(3), R13(2)}	{{1, 2}, {}, {{3, 4}}}	(1, 1)
7	{R8(3), R9(2)}	{{3}, {4}, {{1, 2}}}	(1, 1)
8	{R2(2), R7(3)}	{{4}, {3}, {{1, 2}}}	(1, 1)
9	{R8(3), R12(2)}	{{3}, {1}, {{2, 4}}}	(1, 1)
10	{R4(2), R7(2), R10(1)}	{{4}, {}, {{1, 2, 3}}}	(1, 2, 1)
11	{R3(2), R6(1), R8(2)}	{{2}, {}, {{1, 3, 4}}}	(1, 1, 1)
12	{R2(2), R5(3), R12(2)}	{{3}, {}, {{1, 2, 4}}}	(1, 2, 1)
13	{R3(2), R5(4), R7(3)}	{{1}, {}, {{2, 3, 4}}}	(1, 1, 1)
14	{R5(4), R7(5), R11(2)}	{{1}, {}, {{2, 3, 4}}}	(1, 1, 1)
15	{R3(5), R5(4), R12(3)}	{{1}, {}, {{2, 3, 4}}}	(1, 1, 1)
16	{R3(6), R7(3), R14(4)}	{{1}, {}, {{2, 3, 4}}}	(1, 1, 1)
17	{R5(6), R7(5), R10(2)}	{{}, {}, {{1, 2, 3, 4}}}	(1, 1, 1)
18	{R7(5), R9(4), R11(6)}	{{1}, {}, {{2, 3, 4}}}	(1, 1, 1)
19	{R3(9), R12(3), R14(4)}	{{1}, {}, {{2, 3, 4}}}	(1, 1, 1)
20	{R7(9), R11(6), R14(4)}	{{1}, {}, {{2, 3, 4}}}	(1, 1, 1)

Table V: List of MASs for 1-constituent CCRN $L = 4$.

-
- [1] Tomas Veloz, Pablo Razeto-Barry, Peter Dittrich, and Alejandro Fajardo. Reaction networks and evolutionary game theory. *Journal of mathematical biology*, 68:181–206, 2014.
- [2] Eric Smith and Harold J Morowitz. *The origin and nature of life on earth: the emergence of the fourth geosphere*. Cambridge University Press, 2016.
- [3] Irving R Epstein and John A Pojman. *An introduction to nonlinear chemical dynamics: oscillations, waves, patterns, and chaos*. Oxford university press, 1998.
- [4] Peter Schuster. What is special about autocatalysis? *Monatshefte für Chemie-Chemical Monthly*, 150:763–775, 2019.
- [5] Wim Hordijk and Mike Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of theoretical biology*, 227(4):451–461, 2004.
- [6] Jakob L Andersen, Christoph Flamm, Daniel Merkle, and Peter F Stadler. Defining autocatalysis in chemical reaction networks. *arXiv preprint arXiv:2107.03086*, 2021.
- [7] Zhen Peng, Alex M Plum, Praful Gagrani, and David A Baum. An ecological framework for the analysis of prebiotic chemical reaction networks. *Journal of theoretical biology*, 507:110451, 2020.
- [8] Nicola Vassena and Peter F Stadler. Unstable cores are the source of instability in chemical reaction networks. *arXiv preprint arXiv:2308.11486*, 2023.
- [9] Abhishek Deshpande and Manoj Gopalkrishnan. Autocatalysis in reaction networks. *Bulletin of mathematical biology*, 76(10):2570–2595, 2014.

- [10] Martin Feinberg. Necessary and sufficient conditions for detailed balancing in mass action systems of arbitrary complexity. *Chemical Engineering Science*, 44(9):1819–1827, 1989.
- [11] Alex Blokhuis, David Lacoste, and Philippe Nghe. Universal motifs and the diversity of autocatalytic systems. *Proceedings of the National Academy of Sciences*, 117(41):25230–25236, 2020.
- [12] Fritz Horn and Roy Jackson. General mass action kinetics. *Archive for rational mechanics and analysis*, 47(2):81–116, 1972.
- [13] Martin Feinberg. Foundations of chemical reaction network theory. 2019.
- [14] Polly Y Yu and Gheorghe Craciun. Mathematical analysis of chemical reaction systems. *Israel Journal of Chemistry*, 58(6-7):733–741, 2018.
- [15] Wilhelm Ostwald. Über autokatalyse. *Berichte über die Verhandlungen der Königlich-Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse*, 42:189–191, 1890.
- [16] Zhen Peng, Klaus Paschek, and Joana C Xavier. What wilhelm ostwald meant by “autokatalyse” and its significance to origins-of-life research: Facilitating the search for chemical pathways underlying abiogenesis by reviving ostwald’s thought that reactants may also be autocatalysts. *BioEssays*, 44(9):2200098, 2022.
- [17] Jakob L. Andersen, Christoph Flamm, Daniel Merkle, and Peter F. Stadler. Chemical transformation motifs—modelling pathways as integer hyperflows. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 16(2):510–523, 2019.
- [18] Eric Smith and Supriya Krishnamurthy. Flows, scaling, and the control of moment hierarchies for stochastic chemical reaction networks. *Physical Review E*, 96(6):062102, 2017.
- [19] Praful Gagrani and Eric Smith. Action functional gradient descent algorithm for estimating escape paths in stochastic chemical reaction networks. *Physical Review E*, 107(3):034305, 2023.
- [20] Gheorghe Craciun. Polynomial dynamical systems, reaction networks, and toric differential inclusions. *SIAM Journal on Applied Algebra and Geometry*, 3(1):87–106, 2019.
- [21] GAM King. Autocatalysis. *Chemical Society Reviews*, 7(2):297–316, 1978.
- [22] Charles R. Johnson, Ronald L. Smith, and Michael J. Tsatsomeros. *Matrix Positivity*. Cambridge Tracts in Mathematics. Cambridge University Press, 2020.
- [23] Manoj Gopalkrishnan. Catalysis in reaction networks. *Bulletin of mathematical biology*, 73(12):2962–2982, 2011.
- [24] Uri Barenholz, Dan Davidi, Ed Reznik, Yinon Bar-On, Niv Antonovsky, Elad Noor, and Ron Milo. Design principles of autocatalytic cycles constrain enzyme kinetics and force low substrate saturation at flux branch points. *Elife*, 6:e20667, 2017.
- [25] Zhen Peng, Jeff Linderth, and David A Baum. The hierarchical organization of autocatalytic reaction networks and its relevance to the origin of life. *PLOS Computational Biology*, 18(9):e1010498, 2022.
- [26] Francesco Avanzini, Nahuel Freitas, and Massimiliano Esposito. Circuit theory for chemical reaction networks. *arXiv preprint arXiv:2210.08035*, 2022.
- [27] Personal conversations with Philippe Nghe.
- [28] Frank P Kelly. *Reversibility and stochastic networks*. Cambridge University Press, 2011.
- [29] Yu Liu and David JT Sumpter. Mathematical modeling reveals spontaneous emergence of self-replication in chemical reaction systems. *Journal of Biological Chemistry*, 293(49):18854–18863, 2018.
- [30] Martin Feinberg. Chemical reaction network structure and the stability of complex isothermal reactors—ii. multiple steady states for networks of deficiency one. *Chemical Engineering Science*, 43(1):1–25, 1988.
- [31] Michael R Garey and David S Johnson. *Computers and intractability*, volume 174. Freeman San Francisco, 1979.
- [32] Wolfgang Banzhaf and Lidia Yamamoto. *Artificial chemistries*. MIT Press, 2015.
- [33] Devlin Moyer, Alan R Pacheco, David B Bernstein, and Daniel Segrè. Stoichiometric modeling of artificial string chemistries. *bioRxiv*, 2020.
- [34] Jakob L Andersen, Christoph Flamm, Daniel Merkle, and Peter F Stadler. A software package for chemically inspired graph transformation. In *Graph Transformation: 9th International Conference, ICGT 2016, in Memory of Hartmut Ehrig, Held as Part of STAF 2016, Vienna, Austria, July 5-6, 2016, Proceedings 9*, pages 73–88. Springer, 2016.
- [35] David F Anderson and Tung D Nguyen. Prevalence of deficiency-zero reaction networks in an erdős-rényi framework. *Journal of Applied Probability*, 59(2):384–398, 2022.
- [36] Angel Garcia-Chung, Marisol Bermúdez-Montaña, Peter F Stadler, Jürgen Jost, and Guillermo Restrepo. Chemically inspired erd\h{o}sr\'enyi oriented hypergraphs. *arXiv preprint arXiv:2309.06351*, 2023.
- [37] David A Baum, Zhen Peng, Emily Dolson, Eric Smith, Alex M Plum, and Praful Gagrani. The ecology–evolution continuum and the origin of life. *Journal of the Royal Society Interface*, 20(208):20230346, 2023.
- [38] Victor Blanco and Praful Gagrani. Python Codes for the Generation of Minimal Autocatalytic Subnetworks, 2 2023. Available at <https://github.com/github/autocatatalyticsubnetworks>.
- [39] AA Balandin. Structural algebra in chemistry. *Acta Physicochim*, 1940.
- [40] Allen Hatcher. *Algebraic topology*. 2005.
- [41] Mark I Dykman, Eugenia Mori, John Ross, and PM Hunt. Large fluctuations and optimal paths in chemical kinetics. *The Journal of chemical physics*, 100(8):5735–5750, 1994.
- [42] Christos H Papadimitriou. On the complexity of integer programming. *Journal of the ACM (JACM)*, 28(4):765–768, 1981.
- [43] Hamdy A Taha. *Integer programming: theory, applications, and computations*. Academic Press, 2014.
- [44] Jakob L Andersen, Christoph Flamm, Daniel Merkle, and Peter F Stadler. Maximizing output and recognizing autocatalysis in chemical reaction networks is np-complete. *Journal of Systems Chemistry*, 3(1):1–9, 2012.

MAS id	Reactions (Flow)	FWMC Partition	Production
1	{R2(1), R15(1)}	{{3}, {}, {{1, 2}}	(1, 1)
2	{R23(1), R25(1)}	{{2, 5}, {}, {{3, 4}}	(1, 1)
3	{R15(1), R23(1)}	{{1, 4}, {}, {{2, 3}}	(1, 1)
4	{R13(3), R25(2)}	{{1, 5}, {2}, {{3, 4}}	(1, 1)
5	{R22(3), R23(2)}	{{4}, {5}, {{2, 3}}	(1, 1)
6	{R2(2), R13(3)}	{{4}, {3}, {{1, 2}}	(1, 1)
7	{R12(3), R25(2)}	{{2, 3}, {1}, {{4, 5}}	(1, 1)
8	{R7(3), R25(2)}	{{1, 3}, {}, {{4, 5}}	(1, 1)
9	{R9(2), R22(3)}	{{1, 4}, {2}, {{3, 5}}	(1, 1)
10	{R6(3), R25(2)}	{{5}, {1}, {{3, 4}}	(1, 1)
11	{R14(3), R15(2)}	{{3}, {4}, {{1, 2}}	(1, 1)
12	{R5(3), R23(2)}	{{1, 2}, {}, {{3, 4}}	(1, 1)
13	{R14(3), R23(2)}	{{2}, {1}, {{3, 4}}	(1, 1)
14	{R15(2), R21(3)}	{{1, 5}, {4}, {{2, 3}}	(1, 1)
15	{R22(3), R25(2)}	{{3}, {2}, {{4, 5}}	(1, 1)
16	{R12(3), R15(2)}	{{3, 4}, {5}, {{1, 2}}	(1, 1)
17	{R14(3), R18(2)}	{{3}, {1}, {{2, 4}}	(1, 1)
18	{R2(2), R11(3)}	{{5}, {4}, {{1, 2}}	(1, 1)
19	{R4(3), R9(2)}	{{5}, {2}, {{1, 3}}	(1, 1)
20	{R9(2), R14(3)}	{{2, 5}, {4}, {{1, 3}}	(1, 1)
21	{R18(2), R21(3)}	{{5}, {3}, {{2, 4}}	(1, 1)
22	{R3(3), R15(2)}	{{1}, {}, {{2, 3}}	(1, 1)
23	{R4(3), R23(2)}	{{4}, {1}, {{2, 3}}	(1, 1)
24	{R9(2), R19(3)}	{{1, 2}, {}, {{3, 5}}	(1, 1)
25	{R11(1), R14(2), R22(2)}	{{3}, {}, {{1, 2, 4, 5}}	(1, 1, 1)
26	{R13(2), R21(2), R24(1)}	{{1, 5}, {}, {{2, 3, 4}}	(1, 2, 1)
27	{R11(2), R19(2), R22(1)}	{{1, 3}, {}, {{2, 4, 5}}	(1, 1, 1)
28	{R3(2), R6(1), R14(2)}	{{2}, {}, {{1, 3, 4}}	(1, 1, 1)
29	{R12(2), R20(2), R22(1)}	{{4}, {1}, {{2, 3, 5}}	(1, 1, 1)
30	{R4(2), R13(1), R21(2)}	{{5}, {}, {{1, 2, 3, 4}}	(1, 1, 1)
31	{R6(2), R11(2), R14(1)}	{{5}, {}, {{2, 1, 3, 4}}	(1, 1, 1)
32	{R3(2), R20(1), R22(2)}	{{1, 4}, {}, {{2, 3, 5}}	(1, 1, 1)
33	{R4(1), R12(2), R13(2)}	{{4}, {5}, {{1, 2, 3}}	(1, 1, 1)
34	{R4(2), R13(2), R16(1)}	{{4}, {}, {{1, 2, 3}}	(1, 2, 1)
35	{R5(2), R8(1), R12(2)}	{{2, 3}, {}, {{1, 4, 5}}	(1, 1, 1)
36	{R5(2), R13(2), R21(1)}	{{1, 5}, {}, {{2, 3, 4}}	(1, 1, 1)
37	{R11(2), R18(1), R19(3)}	{{1, 3}, {}, {{2, 4, 5}}	(1, 1, 1)
38	{R19(3), R21(2), R23(1)}	{{2}, {}, {{3, 4, 5}}	(1, 1, 1)
39	{R3(3), R18(2), R25(2)}	{{1, 5}, {}, {{2, 3, 4}}	(1, 1, 2)
40	{R7(2), R19(2), R21(3)}	{{1, 2}, {}, {{3, 4, 5}}	(1, 1, 1)
41	{R8(2), R14(2), R22(3)}	{{3}, {1}, {{2, 4, 5}}	(1, 1, 1)
42	{R12(2), R19(2), R21(3)}	{{2}, {1}, {{3, 4, 5}}	(1, 1, 1)
43	{R11(2), R17(2), R22(3)}	{{1, 3}, {}, {{2, 4, 5}}	(1, 1, 1)
44	{R5(2), R12(3), R21(2)}	{{2}, {}, {{3, 1, 4, 5}}	(1, 1, 1)
45	{R18(2), R19(3), R25(2)}	{{3}, {}, {{2, 4, 5}}	(1, 2, 1)
46	{R12(3), R13(2), R20(2)}	{{4}, {3}, {{1, 2, 5}}	(1, 1, 1)
47	{R19(3), R21(3), R26(1)}	{{2}, {}, {{3, 4, 5}}	(1, 1, 1)
48	{R7(2), R11(3), R19(2)}	{{1, 3}, {}, {{2, 4, 5}}	(1, 1, 1)
49	{R2(2), R5(3), R18(2)}	{{3}, {}, {{1, 2, 4}}	(1, 2, 1)
50	{R16(2), R20(2), R22(3)}	{{4}, {1}, {{2, 3, 5}}	(1, 1, 1)
51	{R4(2), R6(3), R11(4)}	{{5}, {2}, {{1, 3, 4}}	(1, 1, 1)
52	{R14(4), R20(2), R22(3)}	{{}, {1}, {{3, 2, 4, 5}}	(1, 1, 1)
53	{R4(2), R12(4), R20(3)}	{{4}, {1}, {{2, 3, 5}}	(1, 1, 1)
54	{R7(4), R11(3), R14(2)}	{{3}, {}, {{1, 2, 4, 5}}	(1, 1, 1)
55	{R11(3), R13(2), R19(4)}	{{1}, {}, {{3, 2, 4, 5}}	(1, 1, 1)
56	{R13(2), R19(4), R21(3)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1)
57	{R12(4), R14(2), R21(3)}	{{2}, {1}, {{3, 4, 5}}	(1, 1, 1)
58	{R6(4), R11(3), R21(2)}	{{5}, {3}, {{1, 2, 4}}	(1, 1, 1)
59	{R8(4), R19(2), R22(3)}	{{3}, {1}, {{2, 4, 5}}	(1, 1, 1)
60	{R7(4), R14(2), R21(3)}	{{2}, {}, {{1, 3, 4, 5}}	(1, 1, 1)
61	{R4(4), R6(2), R21(3)}	{{5}, {1}, {{2, 3, 4}}	(1, 1, 1)
62	{R5(2), R7(4), R21(3)}	{{1, 2}, {}, {{3, 4, 5}}	(1, 1, 1)
63	{R3(4), R11(2), R22(3)}	{{1}, {}, {{4, 2, 3, 5}}	(1, 1, 1)
64	{R5(2), R14(3), R22(4)}	{{3}, {5}, {{1, 2, 4}}	(1, 1, 1)
65	{R4(3), R11(2), R21(4)}	{{5}, {4}, {{1, 2, 3}}	(1, 1, 1)
66	{R3(2), R5(4), R13(3)}	{{1}, {}, {{2, 3, 4}}	(1, 1, 1)
67	{R8(2), R12(3), R13(4)}	{{}, {3}, {{4, 1, 2, 5}}	(1, 1, 1)
68	{R6(2), R19(4), R21(3)}	{{2}, {1}, {{3, 4, 5}}	(1, 1, 1)
69	{R3(4), R12(2), R14(3)}	{{2}, {5}, {{1, 3, 4}}	(1, 1, 1)
70	{R5(2), R8(3), R22(4)}	{{3}, {2}, {{1, 4, 5}}	(1, 1, 1)
71	{R7(2), R12(3), R20(4)}	{{4}, {3}, {{1, 2, 5}}	(1, 1, 1)
72	{R12(4), R13(3), R22(2)}	{{4}, {5}, {{1, 2, 3}}	(1, 1, 1)
73	{R3(4), R14(5), R26(2)}	{{2}, {5}, {{1, 3, 4}}	(1, 1, 1)
74	{R11(5), R19(4), R26(2)}	{{1}, {}, {{3, 2, 4, 5}}	(1, 1, 1)
75	{R8(4), R17(2), R22(5)}	{{3}, {1}, {{2, 4, 5}}	(1, 1, 1)
76	{R10(2), R12(4), R13(5)}	{{4}, {5}, {{1, 2, 3}}	(1, 1, 1)
77	{R10(2), R12(4), R20(5)}	{{4}, {1}, {{2, 3, 5}}	(1, 1, 1)

MAS id	Reactions (Flow)	FWMC Partition	Production
78	{R1(2), R12(5), R20(4)}	{{4}, {3}, {{1, 2, 5}}	(1, 1, 1)
79	{R7(4), R10(2), R21(5)}	{{2}, {}, {{1, 3, 4, 5}}	(1, 1, 1)
80	{R1(2), R14(5), R22(4)}	{{3}, {5}, {{1, 2, 4}}	(1, 1, 1)
81	{R5(4), R13(5), R17(2)}	{{1}, {}, {{2, 3, 4}}	(1, 1, 1)
82	{R6(4), R11(5), R16(2)}	{{5}, {3}, {{1, 2, 4}}	(1, 1, 1)
83	{R6(5), R11(4), R24(2)}	{{5}, {2}, {{1, 3, 4}}	(1, 1, 1)
84	{R10(2), R12(4), R21(5)}	{{2}, {1}, {{3, 4, 5}}	(1, 1, 1)
85	{R2(3), R9(5), R24(4)}	{{5}, {4}, {{1, 2, 3}}	(1, 1, 2)
86	{R5(5), R9(3), R12(4)}	{{2}, {}, {{1, 3, 4, 5}}	(1, 1, 1)
87	{R2(3), R5(5), R22(4)}	{{3}, {5}, {{1, 2, 4}}	(1, 1, 1)
88	{R2(3), R7(5), R20(4)}	{{4}, {3}, {{1, 2, 5}}	(1, 1, 1)
89	{R8(4), R15(3), R19(5)}	{{3}, {4}, {{1, 2, 5}}	(1, 1, 1)
90	{R11(4), R19(5), R23(3)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1)
91	{R5(5), R7(4), R9(3)}	{{1}, {}, {{3, 4, 5}}	(1, 1, 1)
92	{R3(5), R5(4), R18(3)}	{{1}, {}, {{2, 3, 4}}	(1, 1, 1)
93	{R16(4), R18(5), R25(3)}	{{5}, {1}, {{2, 3, 4}}	(2, 1, 1)
94	{R8(4), R19(5), R23(3)}	{{2}, {1}, {{3, 4, 5}}	(1, 1, 1)
95	{R8(4), R18(3), R19(5)}	{{3}, {1}, {{2, 4, 5}}	(1, 1, 1)
96	{R6(6), R11(5), R17(2)}	{{5}, {3}, {{1, 2, 4}}	(1, 1, 1)
97	{R7(6), R11(5), R17(2)}	{{1}, {}, {{2, 4, 5}}	(1, 1, 1)
98	{R13(3), R16(4), R22(6)}	{{4}, {5}, {{1, 2, 3}}	(1, 1, 1)
99	{R4(6), R21(5), R26(2)}	{{1}, {1}, {{5, 2, 3, 4}}	(1, 1, 1)
100	{R1(2), R4(5), R21(6)}	{{5}, {4}, {{1, 2, 3}}	(1, 1, 1)
101	{R3(6), R13(3), R24(4)}	{{1}, {}, {{2, 3, 4}}	(1, 1, 1)
102	{R5(6), R13(5), R16(2)}	{{}, {}, {{1, 2, 3, 4}}	(1, 1, 1)
103	{R12(6), R21(5), R24(2)}	{{1}, {1}, {{2, 3, 4, 5}}	(1, 1, 1)
104	{R6(5), R10(2), R11(6)}	{{}, {2}, {{5, 1, 3, 4}}	(1, 1, 1)
105	{R12(6), R20(5), R24(2)}	{{1}, {1}, {{4, 2, 3, 5}}	(1, 1, 1)
106	{R5(6), R21(3), R26(4)}	{{1, 2}, {}, {{3, 4, 5}}	(1, 1, 1)
107	{R12(6), R13(5), R24(2)}	{{}, {5}, {{4, 1, 2, 3}}	(1, 1, 1)
108	{R4(3), R11(6), R16(4)}	{{5}, {4}, {{1, 2, 3}}	(1, 1, 1)
109	{R6(6), R21(3), R24(4)}	{{5}, {1}, {{2, 3, 4}}	(1, 1, 1)
110	{R7(6), R21(5), R24(2)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1)
111	{R14(6), R21(3), R26(4)}	{{2}, {1}, {{3, 4, 5}}	(1, 1, 1)
112	{R11(5), R16(6), R18(4)}	{{5}, {3}, {{1, 2, 4}}	(1, 1, 1)
113	{R10(6), R21(5), R23(4)}	{{2}, {1}, {{3, 4, 5}}	(1, 1, 1)
114	{R6(5), R9(4), R24(6)}	{{5}, {2}, {{1, 3, 4}}	(1, 1, 1)
115	{R13(5), R15(4), R17(6)}	{{1}, {}, {{2, 3, 4}}	(1, 1, 1)
116	{R10(6), R13(5), R23(4)}	{{4}, {5}, {{1, 2, 3}}	(1, 1, 1)
117	{R10(6), R20(5), R23(4)}	{{4}, {1}, {{2, 3, 5}}	(1, 1, 1)
118	{R5(9), R9(3), R26(4)}	{{1}, {}, {{3, 4, 5}}	(1, 1, 1)
119	{R3(9), R18(3), R24(4)}	{{1}, {}, {{2, 3, 4}}	(1, 1, 1)
120	{R10(6), R13(9), R16(4)}	{{4}, {5}, {{1, 2, 3}}	(1, 1, 1)
121	{R10(6), R21(9), R26(4)}	{{2}, {1}, {{3, 4, 5}}	(1, 1, 1)
122	{R13(9), R17(6), R24(4)}	{{1}, {}, {{2, 3, 4}}	(1, 1, 1)
123	{R2(1), R9(1), R18(1), R25(1)}	{{5}, {}, {{1, 2, 3, 4}}	(1, 1, 1, 1)
124	{R3(3), R8(3), R12(2), R19(2)}	{{2}, {}, {{1, 3, 4, 5}}	(2, 1, 1, 1)
125	{R9(2), R16(3), R18(2), R24(3)}	{{5}, {}, {{1, 2, 3, 4}}	(1, 1, 1, 1)
126	{R4(3), R7(3), R16(2), R20(2)}	{{4}, {}, {{1, 2, 3, 5}}	(2, 1, 1, 1)
127	{R3(5), R8(3), R10(2), R12(2)}	{{2}, {}, {{1, 3, 4, 5}}	(2, 1, 1, 1)
128	{R3(3), R6(2), R8(3), R19(4)}	{{2}, {}, {{1, 3, 4, 5}}	(2, 1, 1, 1)
129	{R7(3), R10(3), R16(2), R20(5)}	{{3}, {}, {{1, 2, 3, 5}}	(2, 1, 1, 1)
130	{R3(4), R6(2), R10(4), R21(3)}	{{2}, {}, {{1, 3, 4, 5}}	(2, 1, 1, 1)
131	{R3(5), R8(3), R10(4), R23(2)}	{{2}, {}, {{1, 3, 4, 5}}	(2, 1, 1, 1)
132	{R3(3), R5(2), R7(5), R11(4)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
133	{R3(3), R8(5), R19(4), R26(2)}	{{2}, {}, {{1, 3, 4, 5}}	(2, 1, 1, 1)
134	{R3(5), R8(3), R12(4), R24(2)}	{{}, {}, {{2, 1, 3, 4, 5}}	(2, 1, 1, 1)
135	{R3(5), R7(5), R11(4), R24(2)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
136	{R3(7), R6(2), R8(3), R10(4)}	{{2}, {}, {{1, 3, 4, 5}}	(2, 1, 1, 1)
137	{R3(3), R5(6), R7(5), R20(4)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
138	{R3(5), R7(5), R10(2), R11(6)}	{{}, {}, {{1, 2, 3, 4, 5}}	(1, 1, 1, 1)
139	{R3(7), R8(5), R10(4), R26(2)}	{{2}, {}, {{1, 3, 4, 5}}	(2, 1, 1, 1)
140	{R3(3), R5(7), R11(4), R26(5)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
141	{R5(2), R7(8), R11(7), R16(3)}	{{}, {}, {{1, 2, 3, 4, 5}}	(1, 1, 1, 1)
142	{R3(5), R7(5), R9(4), R24(6)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
143	{R7(5), R9(4), R15(7), R17(6)}	{{1}, {}, {{2, 3, 4, 5}}	(2, 1, 1, 1)
144	{R3(3), R5(11), R20(4), R26(5)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
145	{R5(6), R7(8), R17(3), R20(7)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
146	{R3(9), R7(5), R20(4), R24(6)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
147	{R5(7), R11(7), R17(3), R26(8)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
148	{R3(10), R11(4), R24(7), R26(5)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
149	{R7(10), R11(9), R16(5), R24(2)}	{{}, {}, {{1, 2, 3, 4, 5}}	(1, 1, 1, 1)
150	{R7(9), R9(8), R17(3), R24(7)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 2, 1, 1)
151	{R5(9), R7(8), R16(3), R20(7)}	{{}, {}, {{1, 2, 3, 4, 5}}	(1, 1, 1, 1)
152	{R7(10), R10(2), R11(11), R16(5)}	{{}, {}, {{1, 2, 3, 4, 5}}	(1, 1, 1, 1)
153	{R5(10), R11(7), R16(3), R26(8)}	{{}, {}, {{1, 2, 3, 4, 5}}	(1, 1, 1, 1)
154	{R3(10), R10(7), R11(6), R13(5)}	{{}, {}, {{1, 2, 3, 4, 5}}	(1, 1, 1, 1)
155	{R3(10), R9(4), R24(11), R26(5)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
156	{R7(8), R15(6), R17(9), R20(7)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
157	{R9(4), R15(12), R17(11), R26(5)}	{{1}, {}, {{2, 3, 4, 5}}	(2, 1, 1, 1)
158	{R11(7), R15(7), R17(10), R26(8)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
159	{R5(14), R17(3), R20(7), R26(8)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
160	{R3(15), R10(7), R11(6), R18(5)}	{{}, {}, {{1, 2, 3, 4, 5}}	(1, 1, 1, 1)
161	{R3(10), R10(7), R11(11), R26(5)}	{{}, {}, {{1, 2, 3, 4, 5}}	(1, 1, 1, 1)
162	{R3(14), R20(4), R24(11), R26(5)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
163	{R7(10), R9(9), R16(5), R24(11)}	{{}, {}, {{1, 2, 3, 4, 5}}	(1, 1, 1, 1)
164	{R5(17), R16(3), R20(7), R26(8)}	{{}, {}, {{1, 2, 3, 4, 5}}	(1, 1, 1, 1)
165	{R10(7), R11(6), R13(15), R17(10)}	{{}, {}, {{1, 2, 3, 4, 5}}	(1, 1, 1, 1)
166	{R7(14), R17(9), R20(13), R24(6)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
167	{R11(14), R17(10), R24(7), R26(15)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
168	{R15(14), R17(17), R20(7), R26(8)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)
169	{R10(7), R11(21), R17(10), R26(15)}	{{}, {}, {{1, 2, 3, 4, 5}}	(1, 1, 1, 1)
170	{R9(14), R17(10), R24(21), R26(15)}	{{1}, {}, {{2, 3, 4, 5}}	(1, 1, 1, 1)

Table VI: List of MASs for 1-constituent CCRN $L = 5$.