Species subsets and embedded networks of S-systems

Honeylou F. Farinas\textsuperscript{a,b}, Eduardo R. Mendoza\textsuperscript{b,c,d,e}, Angelyn R. Lao\textsuperscript{b,}\textsuperscript{*}

\textsuperscript{a}Department of Mathematics, Mariano Marcos State University, Ilocos Norte, 2906 Philippines
\textsuperscript{b}Mathematics and Statistics Department, De La Salle University, Manila, 0922 Philippines
\textsuperscript{c}Institute of Mathematical Sciences and Physics, University of the Philippines Los Baños, Laguna, 4031 Philippines
\textsuperscript{d}Max Planck Institute of Biochemistry, Martinsried, Munich, Germany
\textsuperscript{e}Faculty of Physics, Ludwig Maximilians University, Geschwister -Scholl- Platz 1, 80539 Munich Germany

Abstract

Magombedze and Mulder (2013) studied the gene regulatory system of \textit{Mycobacterium Tuberculosis} (\textit{Mtb}) by partitioning this into three subsystems based on putative gene function and role in dormancy/latency development. Each subsystem, in the form of \textit{S}-system, is represented by an embedded chemical reaction network (CRN), defined by a species subset and a reaction subset induced by the set of digraph vertices of the subsystem. Based on the network decomposition theory initiated by Feinberg in 1987, we have introduced the concept of incidence-independent and developed the theory of \textit{C}- and \textit{C}-decompositions including their structure theorems in terms of linkage classes. With the \textit{S}-system CRN \textit{N} of Magombedze and Mulder’s \textit{Mtb} model, its reaction set partition induced decomposition of subnetworks that are not CRNs of \textit{S}-system but constitute independent decomposition of \textit{N}. We have also constructed a new \textit{S}-system CRN \textit{N} for which the embedded networks are \textit{C}-decomposition. We have shown that subnetworks of \textit{N} and the embedded networks (subnetworks of \textit{N} ) are digraph homomorphisms. Lastly, we attempted to explore modularity in the context of CRN.

\textbf{Keywords:} Chemical reaction networks, \textit{S}-system, Embedded network, \textit{C}-decomposition, \textit{C}-decomposition, Incidence-independent decomposition,
1. Introduction

S-system models are typically derived from biochemical maps [17], which are based on digraphs whose vertices and arcs represent molecules and interactions of a biochemical system. Modularity of the system is often expressed by partitioning the vertices into subsets and forming corresponding subsystems. In the embedded chemical reaction network (CRN) representation of the S-system [2], each subsystem is represented by an embedded network, defined by a species subset and a reactions subset induced by the set of digraph vertices of the subsystem. In this paper, we study the relationship between the S-system CRN and its embedded networks to gain insight into some properties of the modelled biochemical system.

The paper was motivated by our analysis of two S-system models of the gene regulatory system of *Mycobacterium Tuberculosis* (*Mt*) by Magomedbeze and Mulder [15]. In those models, the vertices (representing genes) were subdivided into three subsets based on putative gene function. We use the S-systems and their subsystems as running examples in the paper.

Our approach is based on the decomposition theory of chemical reaction networks, which was initiated by M. Feinberg in his 1987 review [6]. There, he also introduced the important concept of an independent decomposition: a decomposition \( \{ N_1, ..., N_k \} \) of \( N \) is independent if the stoichiometric subspace of \( N \) is the direct sum of the stoichiometric subspaces of the subnetworks \( N_i \). In [9], it was shown that the deficiency of the network is bounded by the sum of the deficiencies of the subnetworks of an independent decomposition, i.e. \( \delta \leq \delta_1 + ... + \delta_k \).

We introduce the complementary concept of incidence-independence, i.e. when the image of the network’s incidence map is the direct sum of the images of the subnetworks’ incidence maps. The primary example of an incidence-independent decomposition are the linkage classes of a network. For incidence-independent decompositions, \( \delta \geq \delta_1 + ... + \delta_k \), a property familiar from linkage classes. A bi-independent decomposition is one which is both independent and incidence-independent, and for which \( \delta = \delta_1 + ... + \delta_k \).

We observed that the embedded networks of the *Mt* S-systems had the distinctive property that the zero complex was the only complex common to them. We abstracted the concept of a ‘\( C^* \)-decompositions from this: in such a decomposition, the non-zero complexes partition the set of the non-zero complexes of the
network. An interesting subset of $C^*$-decompositions are the $C$-decompositions, where the partition of the reaction set (defining the decomposition) also partitions the set of complexes. Again, the primary example of a $C$-decomposition are the linkage classes. We derive structure theorems for $C$-decompositions and $C^*$-decompositions in terms of linkage classes.

One of our main results is a framework consisting of three components:

- $\mathcal{N}$ and an independent decomposition into $S$-system similar subnetworks
- a new $S$-system CRN $\mathcal{N}^*$ for which the embedded networks are a $C^*$-decomposition
- digraph homomorphisms from subnetworks of $\mathcal{N}$ to the corresponding embedded networks ($\equiv$ subnetworks $\mathcal{N}^*$)

We have organized the contents of the paper as follows:

Section 2 collects the concepts and known results from the theory of chemical reaction networks underlying our results.

Section 3 introduces the NRP (Non-Replicating Phase) and STR (Stationary Phase) models of $\textit{Mtb}$'s gene regulatory system as well as the subsystems of genes $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$.

Section 4 presents a correction of Proposition 10 of [1], which was motivated by the calculation of the deficiency of the NRP CRN, whose value was 39, instead of the predicted 40. The correct formula is an inequality and a new proof using decomposition theory is provided. Incidence-independence of the species decomposition of the $S$-system CRN is a sufficient condition for equality, as in the original proposition.

Section 5 collects and analyzes the properties of the NRP and STR networks as well as those of their embedded networks. A comparison with those from the BST Case studies of [2] show a very high degree of coincidence. Concordance tests of current tools cannot handle the large networks we have been studying. However for $S$-system CRNs, we have found a proof of discordance for all networks with at least two species.

Section 6 develops the theory of $C$- and $C^*$-decompositions including their structure theorems in terms of linkage classes. We then construct $\mathcal{N}^*$, the $*$-disjoint union of embedded networks and show that the embedded networks constitute an independent $C^*$-decomposition of $\mathcal{N}^*$.

In the final results section (Section 7), we complete the previously mentioned framework by introducing the $S$-system similar subnetworks $\mathcal{N}_i$ that constitute an
independent decomposition of \( \mathcal{N} \) as well as the digraph homomorphisms \( \mathcal{N}_i \to \mathcal{N}_i^* \). We then show that the modularity of the decomposition of the \( S \)-system \( \mathcal{N} \) (considered as a digraph division) is not consistent with the modularity of the original digraph model of the gene regulatory system. We introduce the concept of species coupling level to suggest that for a CRN, stoichiometric properties are needed for an appropriate modularity concept.

2. Fundamentals of chemical reaction networks

In this section, we review concepts of subnetworks and embedded networks, decomposition and CRN representation of an \( S \)-system. For a background on Chemical Reaction Network Theory (CRNT), the reader is referred to Appendix. The discussion includes the notations, concepts and results pertinent to this paper.

2.1. Subnetworks and embedded networks of a CRN

A subnetwork and an embedded network of a CRN were introduced by Joshi and Shui in [11] and [10], respectively. The definitions are based on the concept of restriction maps between subsets in a network’s sets.

**Definition 1.** Let \((\mathcal{S}, \mathcal{C}, \mathcal{R})\) be a chemical reaction network. Consider a subset of the species \( S \subset \mathcal{S} \), a subset of the complexes \( C \subset \mathcal{C} \) and a subset of the reactions \( R \subset \mathcal{R} \).

1. The **restriction of \( R \) to \( S \)**, denoted by \( R|_S \), is the set of reactions obtained by taking the reactions in \( R \) and removing all species not in \( S \) from the reactant and product complexes. If a reactant or a product complex does not contain any species from the set \( S \), then the complex is replaced by the 0 complex in \( R|_S \). If a trivial reaction (one in which the reactant and product complexes are the same) is obtained in this process, then that reaction is removed. Also, extra copies of repeated reactions are removed.
2. The **restriction of \( C \) to \( R \)**, denoted by \( C|_R \), is the set of (reactant and product) complexes of the reactions in \( R \).
3. The **restriction of \( S \) to \( C \)**, denoted by \( S|_C \), is the set of species that are in the complexes in \( C \).

**Definition 2.** Let \((\mathcal{S}, \mathcal{C}, \mathcal{R})\) be a chemical reaction network.
1. A subset of the reactions $\mathcal{R}' \subset \mathcal{R}$ defines the **subnetwork** $(\mathcal{S}|_\mathcal{E}|_\mathcal{R}' , \mathcal{C}|_\mathcal{R}' , \mathcal{R}')$, where $\mathcal{C}|_\mathcal{R}'$ denotes the set of complexes that appear in the reactions $\mathcal{R}'$ and $\mathcal{S}|_\mathcal{E}|_\mathcal{R}'$ denotes the set of species that appear in those complexes.

2. An **embedded network** of $(\mathcal{S}, \mathcal{E}, \mathcal{R})$, which is defined by a subset of the species, $S = \{X_{i1}, X_{i2}, \ldots, X_{ik}\} \subset \mathcal{S}$, and a subset of the reactions, $R = \{R_{j1}, R_{j2}, \ldots, R_{jl}\} \subset \mathcal{R}$, that involve all species of $S$, is the network $(S, \mathcal{E}|_{R_{l1}}, R_{l1} )$ consisting of the reactions $R_{l1}$.

In the context of this paper, the subnetworks and embedded networks of the CRN representation of an $S$-system are not necessarily the same. We used $\mathcal{N}_i$ and $\mathcal{N}_i^*$ to denote the subnetwork and embedded network of a CRN $\mathcal{N}$, respectively. The set of subnetworks $\mathcal{N}_i$ is obtained by removing a subset of reactions, while an embedded network $\mathcal{N}_i^*$ is obtained by removing a subset of reactions or subsets of species or both. For instance, removing the species $X_2$ from the reaction $X_1 + X_2 \rightarrow X_1 + X_3$ results in the reaction $X_1 \rightarrow X_1 + X_3$. In some cases, removing species results in a trivial reaction where the source and product complex are identical. For instance, removal of both $X_2$ and $X_3$ from $X_1 + X_2 \rightarrow X_1 + X_3$ results in the trivial reaction $X_1 \rightarrow X_1$. So, after removing species, any trivial reactions and any copies of duplicate reactions are discarded.

**Example 1.** Consider the $S$-system ARL3-S of Arceo et al. in [2] with 4 species $\mathcal{S} = \{X_1, X_2, X_3, X_4\}$ and 8 reactions $\mathcal{R} = \{R_1, R_2, \ldots, R_8\}$ with the following CRN $\mathcal{N}$:

\[
\begin{align*}
R_1 : & \quad X_1 \rightarrow 2X_1 \\
R_2 : & \quad X_1 + X_2 + X_4 \rightarrow X_2 + X_4 \\
R_3 : & \quad X_1 + X_2 \rightarrow X_1 + 2X_2 \\
R_4 : & \quad X_1 + X_2 \rightarrow X_1 \\
R_5 : & \quad X_2 + X_4 \rightarrow X_3 + X_2 + X_4 \\
R_6 : & \quad X_2 \rightarrow) \\
R_7 : & \quad X_2 + X_3 \rightarrow X_3 + X_2 + X_4 \\
R_8 : & \quad X_4 \rightarrow 0
\end{align*}
\]

Suppose that we will form 2 subnetworks ($\mathcal{N}_1$ and $\mathcal{N}_2$) and 2 embedded networks ($\mathcal{N}_1^*$ and $\mathcal{N}_2^*$) of $\mathcal{N}$. To form the subnetworks $\mathcal{N}_1$ and $\mathcal{N}_2$, we simply partition the eight reactions. Suppose we partition the reactions as follows: $\mathcal{R}_{\mathcal{N}_1} = \{R_1, R_2, R_3, R_4\}$ and $\mathcal{R}_{\mathcal{N}_2} = \{R_5, R_6, R_7, R_8\}$. Thus, $\mathcal{N}_1$ and $\mathcal{N}_2$ consist of the set of reactions $\mathcal{R}_{\mathcal{N}_1}$ and $\mathcal{R}_{\mathcal{N}_2}$, respectively. Now, for the embedded networks $\mathcal{N}_1^*$ and $\mathcal{N}_2^*$, we partition the four species. Suppose we have the following partitions: $\mathcal{S}_{\mathcal{N}_1^*} = \{X_1, X_2\}$ and $\mathcal{S}_{\mathcal{N}_2^*} = \{X_3, X_4\}$. Applying (1) of Definition 1, we have the following set of reactions: $\mathcal{R}_{\mathcal{N}_1^*} = \{X_1 \rightarrow 2X_1, X_1 + X_2 \rightarrow X_2, X_1 + X_2 \rightarrow X_1 + 2X_2, X_1 + X_2 \rightarrow X_1\}$ and $\mathcal{R}_{\mathcal{N}_2^*} = \{0 \rightarrow X_3, X_3 \rightarrow 0, X_3 \rightarrow X_3 + X_4, X_4 \rightarrow 0\}$.
2.2. Decomposition of a CRN

Decomposition Theory was initiated by Feinberg in his 1987 review paper. In this paper, we introduced several decompositions as a result of our CRN analysis of Magombedze and Mulder’s [15] Mtb S-systems. The incidence-independent and species decompositions will be discussed in Section 2 and the $C^*$ and $C$-decompositions in Section 6.1. Their relationship with that of Feinberg’s decomposition are shown in Appendix B.

Feinberg’s general concept of a network decomposition is as follows:

**Definition 3.** A set of subnetworks $N_i = (\mathcal{N}_i, \mathcal{C}_i, \mathcal{R}_i)$ is a network decomposition of $N = (\mathcal{N}, \mathcal{C}, \mathcal{R})$ if $\{\mathcal{R}_i\}$ forms a partition of $\mathcal{R}$.

Note that partitioning the reaction set does not necessarily partition the set of complexes. A basic property of decomposition is that $s \leq \sum s_i$ where $s$ and $s_i$ are the dimensions of the stoichiometric subspaces of the network $N$ and its corresponding subnetworks $N_i$, respectively. Meanwhile, if they are equal, then the network is independent [6].

The best known and most studied decomposition is the linkage class decomposition. Independence of linkage classes (ILC) is an important feature of a CRN. Boros [4] showed that a necessary and sufficient condition for a linkage class to be independent is $\delta = \sum \delta_i$. Generally, if a network has an independent decomposition then $\delta \leq \sum \delta_i$ [9].

When considering partitions of the reaction set of a CRN, we also call the decompositions a refinement or coarsening of the decomposition.

**Definition 4.** If $\mathcal{P} = \{\mathcal{P}_i\}$ and $\mathcal{P'} = \{\mathcal{P'}_j\}$ are partitions of a set, then $\mathcal{P}$ is a refinement of $\mathcal{P'}$ if each $\mathcal{P}_i$ is contained in (exactly) one $\mathcal{P'}_j$.

It is easy to show that this property is equivalent to each $\mathcal{P'}_j$ being the disjoint union of some $\mathcal{P}_i$’s. We also say the $\mathcal{P}$ is finer than $\mathcal{P'}$, $\mathcal{P'}$ is coarser than $\mathcal{P}$ and $\mathcal{P'}$ is a coarsening of $\mathcal{P}$.

We have the following elementary but useful Proposition:

**Proposition 1.** If a decomposition is independent, then any coarsening of the decomposition is independent.

**Proof.** Suppose $x$ is in the intersection of the stoichiometric subspaces of two subnetworks of a coarsening. Since each stoichiometric subspace is the direct sum of subspaces from the independent refinement, then the $x$ is the sum of elements from each subnetwork. It follows that $x = 0$. □
2.3. CRN representation of an S-system

An S-system is a set of first-order differential equations that all have the same structure: the derivative of a variable is equal to the difference of two products of power-law functions. The formalism is based on using linear Taylor approximation in logarithmic space. An S-system with $m$ dependent variables $X_i$ is defined as

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^{m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{m} X_j^{h_{ij}}, \ i \in \{1, \ldots, m\}$$

where the dependent variables and the constants $\alpha_i$ and $\beta_i$ are nonnegative, and the exponents $g_{ij}$ and $h_{ij}$ are real.

To enable the use of tools from CRNT in enhancing understanding of S-system dynamics, Arceo et. al introduced in [1] the CRN representations of an S-system using a biochemical map. A biochemical map for such an S-system is the following:

For each dependent variable $X_i$, where $X_{g_{ij}}$ is a variable with $g_{ij} \neq 0$ in the production term and $X_{h_{ij}}$ is one with $h_{ij} \neq 0, i \neq j$ in the degradation term. The CRN representation of this biochemical map is the CRN representation of the S-system. In other words, for each $X_i$, the input reaction is $\sum X_{g_{ij}} \rightarrow X_i + \sum X_{k_{ij}}$ and the output reaction is $X_i + \sum X_{h_{ij}} \rightarrow \sum X_{h_{ij}}$. For each independent variable $X_k$, we add a reaction $0 \rightarrow X_k$. (see Figure [1])

3. The running examples: S-system models of gene regulatory systems of *Mycobacterium tuberculosis*

Magombedze and Mulder [15] developed two models of gene regulatory systems of *Mycobacterium tuberculosis* that used principles of biochemical pathway modelling of systems biology into S-systems. The models correspond to two
phases of the pathogen’s life cycle: the Non-Replicating Phase (NRP) and the Stationary Phase (STR). The NRP model simulates the development of dormant \(Mtb\) by gradual oxygen depletion over 80 days. While the STR model simulates latent \(Mtb\) over 60 days. The gene regulatory system of \(Mtb\) was formalized as digraphs (see Figure 2), with the vertices representing the genes and the arcs representing the interactions between them.

An S-system - with the digraph as its “biochemical map”- is then built for both NRP (left side of Figure 2) and STR (right side of Figure 2) models with the genes (vertices of the digraph) as variables. The models assumed that gene interactions will result in the change of expression of one gene, in such a way that all interactions should contribute to the term which is modelled as a function responsible for gene regulation in the system. Thus, the S-systems for NRP and STR considered the genes as the dependent variables \(X_i\) and no independent variables. It is represented as

\[
\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^{m} X_j^{g_{i,j}} - \beta_i X_i, \quad i \in 1, ..., m
\]

where the constants \(\alpha_i\) and \(\beta_i\) are nonnegative and the exponent \(g_{i,j}\) is real. The S-system ODEs for the NRP and STR models can be found in the Supplementary Information of [15].

A vertex set partition, based on putative gene function and role in dormancy/latency
development, defines three subdigraphs in both models: $D_1$, $D_2$ and $D_3$. Subdiagram $D_1$, is constituted by genes that are involved in the bacilli’s cell wall process. Genes that have functions related to the DosR-regulon and adaptation are clustered in $D_3$ and several genes that have other functions constitute $D_2$. An arc is deleted between a vertex belonging to different subdigraphs. These subdigraphs correspond to the subsystems of the $S$-system for the NRP and STR models.

In this paper, we used CRN to represent the corresponding NRP and STR $S$-systems of [15] and study interesting network properties it exhibit. The CRN representation of each $S$-system is referred to as the “full network” $\mathcal{N}$ with the genes (variables) as species and each subsystem is referred to as the three embedded networks $\mathcal{N}_1^*$, $\mathcal{N}_2^*$ and $\mathcal{N}_3^*$. These embedded networks resulted from the vertex (species) set partition of $\mathcal{N}$.

The CRNs for the NRP and STR models as well as the three embedded networks in each model are documented in the Supplementary Information. Table 1 summarizes the network numbers for the CRN representations of the $S$-system of the NRP and STR full models $\mathcal{N}$ and their corresponding embedded networks $\mathcal{N}_i^*$.

| Network numbers | NRP | \( \mathcal{N}_1^* \) | \( \mathcal{N}_2^* \) | \( \mathcal{N}_3^* \) | STR | \( \mathcal{N}_1^* \) | \( \mathcal{N}_2^* \) | \( \mathcal{N}_3^* \) |
|-----------------|-----|-----------------|-----------------|-----------------|-----|-----------------|-----------------|-----------------|
| $m$             | 40  | 8               | 17              | 15              | 37  | 7               | 14              | 16              |
| $m_{rev}$       | 0   | 0               | 0               | 0               | 1   | 0               | 1               | 1               |
| $n$             | 98  | 19              | 36              | 42              | 95  | 19              | 29              | 44              |
| $n_r$           | 60  | 10              | 20              | 28              | 60  | 12              | 17              | 29              |
| $r$             | 80  | 16              | 34              | 30              | 74  | 14              | 28              | 32              |
| $r_{rev}$       | 80  | 16              | 34              | 30              | 72  | 14              | 26              | 30              |
| $l$             | 19  | 3               | 3               | 12              | 22  | 5               | 2               | 13              |
| $sl$            | 98  | 19              | 36              | 42              | 94  | 19              | 28              | 43              |
| $t$             | 38  | 9               | 16              | 14              | 35  | 7               | 12              | 15              |
| $s$             | 40  | 8               | 17              | 15              | 37  | 7               | 14              | 16              |
| $q$             | 40  | 8               | 17              | 15              | 37  | 7               | 14              | 16              |
| $\delta$        | 39  | 8               | 16              | 15              | 36  | 7               | 13              | 15              |
| $\delta_p$      | 20  | 2               | 3               | 13              | 23  | 5               | 3               | 13              |
4. The deficiency formula for S-system CRNs

In [1], Arceo et al. introduced the concepts of reversibility and non-self regulation of dependent species (variables) of an S-system and derived a formula for calculating the total deficiency of a non-self regulating (NSR) S-systems (Proposition 10 of [1]). It turns out that the equality in the formula does not hold in general. In this Section, we present two counterexamples: the NRP full network $\mathcal{N}$ and its embedded network $\mathcal{N}_2^*$. We restate the result and provide a proof based on decomposition theory.

First, we recall definitions from [1].

**Definition 5.** An S-system is called non-self regulating (NSR) if it is monomolecular and each dependent species $X_i$ does not occur in its production term.

**Remark 1.** The term “monomolecular” refers to the biochemical map or digraph underlying the S-system.

**Definition 6.** Let $R_j$ and $P_j$ be the set of variables regulating the inflow and outflow reactions of the species $X_j$ (= dependent variable) of an S-system, respectively. The species is called reversible if $R_j = P_j$. Otherwise, it is called irreversible. An S-system is called reversible (irreversible) if all its species are reversible (irreversible).

The term reversible species comes from the fact that the reaction pair associated with $X_j$ is $R_j \leftrightarrow X_j + R_j$. More generally, we write $R_j \rightarrow X_j + R_j$ and $X_j + P_j \rightarrow P_j$ for the reaction pair associated with a species.

**Example 2.** Both the NRP and STR full models $\mathcal{N}$ and their embedded networks $\mathcal{N}_i^*$ are non-self regulating S-system. But only the NRP $\mathcal{N}$ and its embedded networks $\mathcal{N}_i^*$ are irreversible. The STR $\mathcal{N}$ and some of its embedded networks $\mathcal{N}_i^*$ have a reversible species.

Next, we state the formula in computing for the total deficiency of an NSR S-system.

**Proposition 2.** (Proposition 10 in [1]). Let $m$ be the number of dependent variables (species) in an NSR S-system. If $m_{re_v}$ is the number of reversible species then the total deficiency is $m - m_{re_v}$. 
As mentioned earlier, the equality in the formula does not hold in general. Computations for the two counterexamples of this proposition show that in NRP $\mathcal{N}$ it has $m = 40$ species (all of which are dependent variables), no reversible ones ($m_{rev} = 0$) but deficiency $\delta = 39 < 40 - 0$. Also, $\delta = 13 < 14 - 0$ in NRP $\mathcal{N}_2^*$ (see Table 1).

A decomposition theory approach is used to restate and provide a new proof of this Proposition. Fundamental concepts and known results of decomposition theory are discussed in Section 2.2. But before presenting the reformulated deficiency formula for $S$-system CRNs, we first introduce the new concept of “incidence-independence” and derive its basic properties. Incidence-independence is a natural complement to the “independence” property of a decomposition, but has so far been neglected in the CRNT literature.

**Definition 7.** A decomposition $\{\mathcal{N}_1, ..., \mathcal{N}_k\}$ of a CRN is **incidence-independent** if and only if the image of the incidence map of $\mathcal{N}$ is the direct sum of the images of the incidence maps of the subnetworks.

Since the direct sum property of the images is equivalent to the dimension of the image of $I_a = \text{the sum of the dimensions of the subnetwork incident map images}$, an equivalent formulation is the following equality:

$$n - l = \sum (n_i - l_i). \quad (1)$$

**Example 3.** The linkage classes form the primary example of an incidence-independent decomposition, since $n = \sum n_i$ and $l = \sum l_i$.

We have the following proposition which is analogous to Proposition 1:

**Proposition 3.** If a decomposition is incidence-independent, then any coarsening of the decomposition is incidence-independent.

The proof has the same argumentation as in Proposition 1 now applied to the image of the incidence map instead of the stoichiometric subspace.

The following proposition is a familiar property of linkage classes:

**Proposition 4.** For an incidence independent decomposition $\mathcal{N} = \mathcal{N}_1 + ... + \mathcal{N}_k$, then $\delta \geq \delta_1 + ... + \delta_k$.

**Proof.** Since for any decomposition, $s \leq \sum s_i$, subtracting the LHS from $n - l$ and the RHS from $\sum n_i - \sum l_i$ delivers the claim. $\square$
Definition 8. A decomposition is bi-independent if it is both independent and incidence-independent.

The ILC property of linkage classes is the best known example of a bi-independent decomposition.

Corollary 1. For a bi-independent decomposition, \( \delta = \delta_1 + \ldots + \delta_k \).

Proof. It was shown in [9], that for an independent decomposition: \( \delta \leq \delta_1 + \ldots + \delta_k \).
Hence claim follows directly from the previous proposition. \( \square \)

Another new concept that emerged in our restatement of the deficiency formula for S-system embedded networks is the species decomposition.

Definition 9. Let \( R_i \) and \( P_i \) be the set of variables regulating the inflow and outflow reactions of the species \( X_i \) (= dependent variable) of an S-system, respectively. A set of subnetworks \( \{ \mathcal{N}_i = (\mathcal{I}_i, \mathcal{C}_i, \mathcal{R}_i), i = 1, \ldots, m \} \) is a species decomposition of an S-system \( \mathcal{N} \) if the subnetwork is formed by the pair of reactions associated with each species, i.e, \( \mathcal{R}_i = \{ R_i \rightarrow X_i + R_i, X_i + P_i \rightarrow P_i \} \).

Remark 2. A species decomposition is another way of decomposing a network since it partitions the reaction set. Also, in this decomposition, a network with \( m \) species results into \( m \) subnetworks.

Example 4. Consider an S-system with 3 species \( \{ X_1, X_2, X_3 \} \). Since each species corresponds to a pair of reactions, we have 6 reactions in total. Suppose we have the following CRN \( \mathcal{M} \): \( \{ X_1 \rightarrow 2X_1, X_1 + X_2 + X_3 \rightarrow X_2 + X_3, X_1 + X_2 \rightarrow X_1 + 2X_2, X_1 + X_2 \rightarrow X_1, X_2 + X_3 \rightarrow 2X_3 + X_2, X_3 \rightarrow 0 \} \). We have the following species decomposition (subnetworks) of \( \mathcal{M} \):
\[
\mathcal{M}_1 = \{ X_1 \rightarrow 2X_1, \quad X_1 + X_2 + X_3 \rightarrow X_2 + X_3 \} \\
\mathcal{M}_2 = \{ X_1 + X_2 \rightarrow X_1 + 2X_2, \quad X_1 + X_2 \rightarrow X_1 \} \\
\mathcal{M}_3 = \{ X_2 + X_3 \rightarrow 2X_3 + X_2, \quad X_3 \rightarrow 0 \}
\]

Finally, we now reformulate the deficiency formula for S-system CRNs. The correct formula is an inequality, instead of an equality.

To avoid the minor complications of independent variables, we consider the embedded representation of an S-system. In particular, the species set \( \mathcal{I}' \) consists only of the dependent variables \( \{ X_1, \ldots, X_m \} \).

Theorem 1. Let \( \mathcal{N}' = (\mathcal{I}', \mathcal{C}', \mathcal{R}') \) be the embedded representation of an S-system and \( \mathcal{R}_i' \) denotes the set \( \{ R_i \rightarrow X_i + R_i, X_i + P_i \rightarrow P_i \} \).
i) The sets \( \{ \mathcal{R}'_i \} \) form a partition of the reaction set \( \mathcal{R}' \) and hence induce a species decomposition.

ii) The species decomposition is independent and implies that \( \delta \leq m - m_{rev} \). If the species decomposition is also incidence-independent, i.e. bi-independent, then \( \delta = m - m_{rev} \).

**Proof.** i) We have to show that if \( i \neq j \), then the intersection of the reaction sets is empty. Suppose that the sets \( \{ \mathcal{R}'_i \} \) do not form a partition of the reaction set \( \mathcal{R}' \). Then there exists two sets \( \mathcal{R}'_i \) and \( \mathcal{R}'_j \), where \( i \neq j \), that has a common reaction. We consider the following cases: a) two inflow reactions coincide and b) an inflow reaction coincides with an outflow reaction. The remaining cases involve converse reactions and hence follow similarly.

We let \( \mathcal{R}'_i = \{ R_i \rightarrow X_i + R_i, X_i + P_i \rightarrow P_i \} \) and \( \mathcal{R}'_j = \{ R_j \rightarrow X_j + R_j, X_j + P_j \rightarrow P_j \} \). We denote the subvectors of \( R_i, R_j, P_i \) and \( P_j \) as \( V_i, V_j, W_i \) and \( W_j \), respectively. We set \( V_i = (a_1, ..., a_m) \) and \( W_j = (b_1, ..., b_m) \). In connection to the \( a \) elements of \( V_i \), the two input reactions in \( \mathcal{R}'_i \) and \( \mathcal{R}'_j \) coincide thus \( R_i = R_j \) and \( X_i + R_i = X_j + R_j \). This implies that \( V_i = V_j \) and \( X_i + V_i = X_j + V_i \) or \( (a_1, ..., a_i + 1, ..., a_m) = (a_1, ..., a_j + 1, ..., a_m) \). Since \( i \neq j \), \( a_i + 1 = a_i \) and \( a_j = a_j + 1 \), a contradiction.

As for the \( b \) elements of \( W_j \), we assume that an inflow reaction in \( \mathcal{R}'_i \) coincides with an \( \mathcal{R}'_j \). Then \( R_i = X_j + P_j \) and \( X_i + R_i = P_j \). Thus, we have \( V_i = X_j + W_j \) or \( (a_1, ..., a_i, ..., a_m) = (b_1, ..., b_j + 1, ..., b_m) \). This implies that \( a_i = b_i \) and \( a_j = b_j + 1 \). Similarly, \( V_i + X_i = W_j \) implies that \( a_i + 1 = b_i \) and \( a_j = b_j \). Since \( i \neq j \), \( a_i = b_i = a_i + 1 \) and \( b_j = a_j = b_j + 1 \), a contradiction.

ii) Note that the stoichiometric subspace \( S_i \) of each of the \( m \) subnetworks \( \mathcal{R}'_i \) of the species decomposition is \( \{ X_i \} \). Thus, the rank of each \( \mathcal{R}'_i \) is 1. Since there are \( m \) subnetworks and the rank of an \( S \)-system is \( m \), \( s = m = s_1 + ... + s_m \) and this implies independence. In a species decomposition, the subnetworks are either reversible or irreversible. Each reversible species generates 2 distinct complexes and 1 linkage class, while an irreversible one generates 4 distinct complexes and 2 linkage classes. Hence, \( \delta_i = 0 \) if \( X_i \) is reversible and \( \delta_j = 1 \) if \( X_j \) is irreversible.

Since the species decomposition is independent, from Proposition 3.3 of [9], we have \( \delta \leq \delta_1 + ... + \delta_m = m - m_{rev} \). From Corollary [1], if the decomposition is incidence-independent too, then equality holds.

**Example 5.** The following table shows some of the network numbers for the species decomposition of \( \mathcal{M} \) in Example [4]

| Example | Network Numbers |
|---------|-----------------|
|         |                 |

The species decomposition of \( \mathcal{M} \) is bi-independent since \( s = \sum s_i \) and \( \delta = \sum \delta_i \). Also, \( \mathcal{N} \) has no reversible species thus, \( \delta = 3 = m - m_{rev} \).
The following corollary provides a sufficient condition for the species decomposition of an $S$-system to be incidence-dependent:

**Corollary 2.** If $\{\mathscr{S}_1, ..., \mathscr{S}_k\}$ is a partition of the species set, then the sets $R(\mathscr{S}_j) := \{R_i \mid X_i \in \mathscr{S}_j\}$ form a partition of the reaction set and hence induce an independent decomposition of the network. The deficiency of the network is less than or equal to the sum of the deficiencies of the subnetworks. If the deficiency is less than the sum, then the species decomposition of the network is incidence-dependent.

**Example 6.** Since for the full network $\mathcal{N}$ of the NRP model, deficiency $\delta = 39 < 40 = m - m_{\text{rev}}$, the species decomposition of the network is incidence-dependent. The species decomposition of the NRP embedded representation $\mathcal{N}^*_2$ is also incidence-dependent since its deficiency $\delta = 14 < 13 = m - m_{\text{rev}}$.

### 5. Properties of the S-system CRNs and their embedded networks

We discuss some important network properties of NRP and STR full networks and their embedded subnetworks. In Section 5.1, we consider their point terminality (PT), sufficient reactant diversity (SRD), terminality boundedness by deficiency (TBD) and equality of reactant and stoichiometric subspaces (RES) as well as the relationships between these characteristics. In Section 5.2, we derive their discordance from a general result about S-system networks.

#### 5.1. Network properties of the NRP and STR models

The NRP and STR full models and their embedded networks are fully open networks since each species has an outflow reaction as seen on their corresponding CRN in the Supplementary Information. Hence, the stoichiometric subspace $S = \mathbb{R}^\mathcal{S}$ thus, the rank of the network is equal to the number of species, i.e $s = m$ as shown in the first column of Tables 2 and 3.
In a weakly reversible (WR) network, $s_l = l$. The network numbers of the NRP and STR models in Table 1 shows that $s_l \neq l$ hence, the networks are not weakly reversible. Since weakly reversible CRNs are $t$-minimal ($t = l$), the “NO’s” in the third column follow from those in the second column of Tables 2 and 3.

We used CoNtRol, an online toolbox to test the endotacticity (ET) of the network. Due to the complexity of some networks, CoNtRol was not able to generate the said test. That is why the ET column in Tables 2 and 3 is incomplete. We also used another software tool, ERNEST, consisting of extensible Matlab code which is specifically valuable for large networks such as the NRP and STR to determine the independent linkage class (ILC) property and the network numbers in Table 1. From the report, it showed that the rank of the networks are not equal to the sum of the rank of their corresponding linkage classes. Thus, the “NO’s” in the ILC column.

The WR, TM and ET columns in Tables 2 and 3 are well-studied network properties in the current CRNT literature. Meanwhile, the point terminal (PT) and terminality bounded by deficiency (TBD) columns were introduced in [2] as additional network properties in their 15 BST case studies. They defined a network to be point terminal if $t = n - n_r$. For the TBD, we first introduce some convenient notation:

**Definition 10.** The **terminality** $\tau(D)$ (or simply $\tau$) of a digraph $D$ is the non-negative integer $t - l$. 

---

**Table 2: NRP Network properties.**

| Model | $s = m$ | WR  | TM  | ET  | ILC | PT  | SRD | TBD | RES |
|-------|--------|-----|-----|-----|-----|-----|-----|-----|-----|
| $N$   | 40     | NO  | NO  | NO  | NO  | YES | YES | YES | YES |
| $N_1^*$ | 8     | NO  | NO  | NO  | NO  | YES | YES | YES | YES |
| $N_2^*$ | 17    | NO  | NO  | NO  | NO  | YES | YES | YES | YES |
| $N_3^*$ | 15    | NO  | NO  | NO  | NO  | YES | YES | YES | YES |

**Table 3: STR Network properties.**

| Model | $s = m$ | WR  | TM  | ET  | ILC | PT  | SRD | TBD | RES |
|-------|--------|-----|-----|-----|-----|-----|-----|-----|-----|
| $N$   | 37     | NO  | NO  | X   | NO  | YES | YES | YES | YES |
| $N_1^*$ | 7     | NO  | NO  | NO  | NO  | YES | YES | YES | YES |
| $N_2^*$ | 14    | NO  | NO  | X   | NO  | YES | YES | YES | YES |
| $N_3^*$ | 16    | NO  | NO  | X   | NO  | YES | YES | YES | YES |
In other words, a \( t \)-minimal digraph is one with zero terminality and a non-\( t \)-minimal one with positive terminality.

**Definition 11.** A CRN is of type **terminality bounded by deficiency** (TBD) if \( t - l \leq \delta \), otherwise, of type **terminality not deficiency-bounded** (TND), i.e. \( t - l > \delta \).

A remark of Feliu and Wiuf [8] that \( q < s \) (i.e. low reactant rank) implied degeneracy of all positive equilibria of a mass action system led us to the concept of “sufficient reactant diversity” (SRD) as a necessary condition for the existence of a non-degenerate equilibrium \( q \geq s \Rightarrow n_r \geq s \).

**Definition 12.** A CRN has **low reactant diversity** (LRD) if \( n_r < s \), otherwise it has **sufficient reactant diversity** (SRD). An SRD network has **high reactant diversity** (HRD) or **medium reactant diversity** (MRD) if \( n_r > s \) or \( = s \), respectively.

The next Proposition clarifies the relationship between deficiency-bounded terminality and sufficient reactant diversity.

**Proposition 5.** Let \( \mathcal{N} \) be a chemical reaction network.

(i) A network with deficiency-bounded terminality (TBD) has sufficient reactant diversity.

(ii) If the network is point terminal, then the converse also holds, i.e. TBD \( \Leftarrow \) SRD (or equivalently TND \( \Rightarrow \) LRD).

**Proof.** From \( \delta - \tau(\mathcal{N}) = n - l - s - (t - l) = n - t - s \), we obtain TBD, i.e. \( \delta \tau(\mathcal{N}) \geq 0 \) iff \( n - t \geq s \). Since \( n_r \geq n - t \) for any network, TBD implies SRD. If \( \mathcal{N} \) is point terminal \( n_r = n - t \), showing that the converse also holds. \( \square \)

**Corollary 3.** A \( t \)-minimal network is an SRD network.

**Remark 3.** We have the following ascending chain of networks: reversible \( \Rightarrow \) weakly reversible \( \Rightarrow \) \( t \)-minimal \( \Rightarrow \) TBD \( \Rightarrow \) SRD.

**Example 7.** Since all the \( S \)-system embedded networks in Tables 2 and 3 are point terminal, Theorem 2(i) tells us that the information in the TBD and SRD columns are the same.
In [3], a classification of CRNs based on the intersection $R \cap S$ of the reactant and stoichiometric subspaces was introduced. Open networks belong to the network class SRS (stoichiometry-determined reactant subspace) with $R \cap S = R$, since $R \subset S = R$. Tables 2 and 3 show that all $S$-system embedded networks belong to the subset of coincident $R$ and $S$ subspaces, which we denote by RES ($R$ Equals $S$).

We observe that, for open networks, $R = S$ is equivalent to $q = s$, i.e. the network’s rank difference $\Delta(N) = 0$.

We include the statements of the Proposition 5 to collect the relationships between TBD, SRD and RES in the following Theorem:

**Theorem 2.**

(i) A TBD network has sufficient reactant diversity. If the network is point terminal, then the converse holds, i.e. $TBD \iff SRD$.

(ii) A TPD network has high reactant diversity (HRD). If the network is point terminal, then the converse also hold, i.e. $TPD \Rightarrow HRD$.

(iii) An RES network has sufficient reactant diversity.

(iv) $RES \cap TBD = RES \iff n_r \geq s + t_c$. If the network is point terminal, then $RES$ implies TBD. If the network is not point terminal, then it has high reactant diversity ($n_r > s$) and reactant deficiency $\delta_\rho > 0$.

**Proof.**

(i) was already shown in Proposition 5.

(ii) $t - l < \delta \Rightarrow s < n - t \Rightarrow s < n_r - t_c \Rightarrow n_r > s$, since $t_c \geq 0$. If the network is point terminal, $t_c = 0$, and the converse holds.

(iii) For any CRN, we have $n_r \geq q$. Since the CRN is RES, $q = s$, which establishes the sufficient reactant diversity.

(iv) $\delta - \tau(N) = n - t - s = n - (t_c + t_p) - s = n - (t_c + n - n_r) - s = n_r - q - t_c = n_r - (s + t_c)$. Hence $TBD \iff n_r \geq s + t_c$, so that the first claim follows. For point terminal networks $t_c = 0$, and (ii) establishes the validity of the RHS. If the network is not point terminal, $t_c > 0$, hence $n_r > s$ and $\delta_\rho := n_r - q = n_r - s > 0$.

**Remark 4.** The inequality $n_r \geq s + t_c$ expresses clearly which additional characteristic an RES network needs beyond SRD to imply deficiency-bounded terminality.

Figure 3 illustrates the relationships for $S$-system embedded networks.

These network properties of the NRP and STR show a very high degree of coincidence with the eight embedded $S$-systems from BST Case Studies in [2] as seen in Table 4.
5.2. The discordance of S-system CRNs

Limitations in the software tools available for Concordance tests in Tables 2 and 3 prevented results regarding this property. However, in this section, we present the discordance of S-system CRNs. Throughout this section, we consider both total and embedded S-system reaction networks and their kinetics.

5.2.1. Concordance, weakly monotonic kinetics and non-inhibitory Power Law Kinetics

We recall from [1] the definition of concordance and weakly monotonic kinetics. We then complete the proof of the basic result relating both concepts to injectivity in non-inhibitory power law kinetics.

In preparation for the definition of concordance, we consider a reaction network \((\mathcal{I}, \mathcal{C}, \mathcal{R})\) with stoichiometric subspace \(S \subset \mathbb{R}^\mathcal{I}\) and we let \(L : \mathbb{R}^\mathcal{R} \to S\) be the linear map defined by

\[
L\alpha = \sum_{y \to y' \in \mathcal{R}} \alpha_{y \to y'}(y' - y).
\]
Note that the real scalar multipliers \( \{ \alpha_{y \to y'} \} \) in Equation (2) are permitted to be positive, negative or zero.

**Definition 13.** The reaction network \((\mathcal{S}, C, \mathcal{R})\) is **concordant** if there do not exist an \( \alpha \in \ker L \) and a nonzero \( \sigma \in S \) having the following properties:

i) For each \( y \to y' \in \mathcal{R} \) such that \( \alpha_{y \to y'} \neq 0 \), supp \( y \) contains a species \( s \) for which \( \sgn \sigma_s = \sgn \alpha_{y \to y'} \).

ii) For each \( y \to y' \in \mathcal{R} \) such that \( \alpha_{y \to y'} = 0 \), either \( \sigma_s = 0 \) for all \( s \in \text{supp } y \) or else supp \( y \) contains a species \( s \) and \( s' \) for which \( \sgn \sigma_s = -\sgn \sigma_{s'} \), both not zero.

A network that is not concordant is **discordant**.

The set of weakly monotonic kinetics is defined as follows:

**Definition 14.** A kinetics \( \mathcal{K} \) for a reaction network \((\mathcal{S}, C, \mathcal{R})\) is **weakly monotonic** if, for each pair of compositions \( c^* \) and \( c^{**} \), the following implications hold for each reaction \( y \to y' \in \mathcal{R} \) such that \( \text{supp } y \subset \text{supp } c^* \) and \( \text{supp } y \subset \text{supp } c^{**} \):

i) \( \mathcal{K}_{y \to y'}(c^{**}) > \mathcal{K}_{y \to y'}(c^*) \Rightarrow \) there is a species \( s \in \text{supp } y \) with \( c^{**}_s > c^*_s \).

ii) \( \mathcal{K}_{y \to y'}(c^{**}) = \mathcal{K}_{y \to y'}(c^*) \Rightarrow \) \( c^{**}_s > c^*_s \) for all \( s \in \text{supp } y \) or else there are species \( s, s' \in \text{supp } y \) with \( c^{**}_s > c^*_s \) and \( c^{**}_{s'} < c^*_{s'} \).

The final concept we need, to formulate the basic result, is that of non-inhibitory Power Law kinetics:

**Definition 15.** A power law kinetics is in \( \mathcal{PL} - \mathcal{NIK}(\mathcal{N}) \) (or has **non-inhibitory kinetics**) if

i) the kinetic order matrix \( F \) is non-negative and

ii) a kinetic order \( f_{r,s} > 0 \) iff the species \( s \) is an element of \( \text{supp } \rho(r) \).

The following theorem of Shinar and Feinberg [16] relates concordance and injectivity in weakly monotonic and non-inhibitory power law kinetics:

**Theorem 3.** For a chemical reaction network \((\mathcal{S}, C, \mathcal{R})\), the following conditions are equivalent:

i) The network is concordant.

ii) The network is injective against all weakly monotonic kinetics.

iii) The network is injective against all non-inhibitory power law kinetics.
Proof. (i)→(ii): This is shown in Proposition 4.8 of [16].
(iii) → (i): Inspection of the proof of Proposition 4.9 (Appendix A of [16]) reveals that the kinetics constructed is actually a non-inhibitory power law kinetics, thus proving this implication.

We now confirm that a non-inhibitory power law kinetics is weakly monotonic, from which (ii) → (iii) follows. To show that a kinetic $\mathcal{K}$ in $\mathcal{PL} - NIK(\mathcal{N})$ fulfills (i) of Definition [14], suppose that for all species in supp $\rho(r)$, $c^*_s \leq c^{**}_s$. Since $f_{r,s} > 0$, then $(c^*_s)^{f_{r,s}} \leq (c^{**}_s)^{f_{r,s}}$. Taking the product on both sides over all $s$ in supp $\rho(r)$, we obtain a contradiction. To show (ii), again suppose there is an $s$ with $c^*_s < c^{**}_s$. Since $f_{r,s} > 0$, then $(c^*_s)^{f_{r,s}} < (c^{**}_s)^{f_{r,s}}$ - this is where property (ii) of the $\mathcal{PL} - NIK$ definition comes in. Unless there is an $s'$ with $c^*_{s'} > c^{**}_{s'}$ and consequently $(c^*_s)^{f_{s,s'}} > (c^{**}_{s})^{f_{s,s'}}$, one cannot obtain the product equality. □

5.2.2. Discordance of $S$-system reaction networks

**Theorem 4.** i) An $S$-system reaction network with 2 or more dependent species is discordant.

ii) A non-constant $S$-system reaction network with one dependent species is discordant if $A \neq 0$ and the product of its rate constants ($\alpha$ and $\beta$) and kinetic orders ($g$ and $h$) $\alpha \beta gh \neq 0$, otherwise concordant.

**Proof.** (i) To show discordance, we only need to construct a $\mathcal{PL} - NIK$ kinetics with a non-injective SFRF on the reaction network. We consider the total representation of the $S$-system first. The SFRF of any PL kinetics maybe decomposed into an $f_D$ (containing all coordinate functions for dependent species) and $f_I$ with the coordinate functions for the independent species. Since each $f_{i,j}$ has the form $f_{i,j} = a_j > 0$, it is clear that the SFRF is injective iff $f_D$ is injective. We will construct a $\mathcal{PL} - NIK$ kinetics whose $f_D$ is multistationary, implying that it is not injective. To achieve this, we choose the kinetic order matrix as follows:

- The kinetics is $\mathcal{PL} - NIK$.
- The first column of the derived $A$ matrix consists only of zeros, i.e. $g_{i1} = h_{i1}$, for all $i = 1, ..., m$.
- $A \neq 0$ (this is possible since the system has 2 or more dependent species).

We have an $S$-system with $\det A = 0$, which according to the above, either has no steady state or infinitely many steady states. Since the rate constants can be
arbitrarily chosen from the positive real numbers, we adjust them so that we have the latter alternative. Since the network is open, this shows that \( f_D \) is not injective in the single stoichiometric class.

(ii) We begin this time with \( A = 0 \). Since we assumed it is not constant, then \( \alpha \neq \beta \). So, the kinetics \( = (\alpha - \beta)X^g = (\alpha - \beta \exp(g\ln X)) \) is strictly monotonic, and hence injective for any positive \( g \). Hence the CRNs are concordant according to Theorem 3. If \( A \) is nonzero, which is equivalent to \( \det A \) nonzero, the system has a unique positive equilibrium. If one of the rate constants or the kinetic orders is 0, then we obtain a function as in the \( A = 0 \) case plus a constant. Hence, the system is also injective. To show that it is not injective for both rate constants and both kinetic orders nonzero, we need to look at its derivative. Since the \( S \)-system in one variable has the form \( f(x) = \alpha X^g - \beta X^h \) and \( f'(x) = \alpha gX^{g-1} - \beta hX^{h-1} \). One easily checks that \( f'(x) < (>)0 \) if and only if \( x^{g-h} < (>)\frac{\beta}{\alpha} \). From Calculus, it follows that the function is strictly decreasing and strictly increasing left and right of the particular \( x \)-value. The value is between its steady state at 0 and its positive steady state at \( \frac{\beta}{\alpha} \), since \( g > h \) (i.e. production \( \beta \), degradation). This means that \( f \) is not injective in the interval \((0, \frac{\beta}{\alpha})\), so that it is discordant according to the Shinar-Feinberg Theorem in the previous section. A similar argument holds for \( h > g \), the function is not injective in the interval \((0, \frac{\beta}{\alpha})\). For the constant case, the same argument as in (i) applies.

These considerations also cover the embedded representation case, since the same \( f_D \) describes the embedded system, by considering the powers of the independent species as part of the rate constants. \( \square \)

6. The \( * \)-disjoint union \( \mathcal{N}^* \) of the embedded networks of \( \mathcal{N} \)

In our study of the embedded networks of the NRP and STR full network, we observed that the zero complex was the only complex common to them. From this, we abstracted the concepts of a \( C^* \)-decomposition and \( C \)-decomposition. In this Section, structure theorems for both decomposition types are derived in terms of linkage classes. Moreover, a new \( S \)-system CRN \( \mathcal{N}^* \) is also constructed for which the embedded networks \( \mathcal{N}_i^* \) form a \( C^* \)-decomposition.

6.1. The \( C^* \)-decomposition and the \( C \)-decomposition

We first formulate a formal definition of a \( C^* \)-decomposition:

**Definition 16.** A decomposition \( \mathcal{N} = \mathcal{N}_1 + \mathcal{N}_2 + \ldots + \mathcal{N}_k \) with \( \mathcal{N}_j = (\mathcal{I}_j, \mathcal{E}_j, \mathcal{R}_j) \) is a \( C^* \)-decomposition if \( \mathcal{E}_i^* \cap \mathcal{E}_j^* = \emptyset \) for \( i \neq j \) where \( \mathcal{E}_i^* \) and \( \mathcal{E}_j^* \) are the non-zero complexes in \( \mathcal{E}_i \) and \( \mathcal{E}_j \), respectively.
The embedded networks $N_i^*$ of the NRP and STR models are $C^*$-decompositions of their full network $N$, respectively.

A particularly interesting subset of $C^*$-decompositions consists of the $C^*$-decomposition.

**Definition 17.** A decomposition $N = N_1 + N_2 + \ldots + N_k$ with $N_i = (S_i, C_i, R_i)$ is a $C^*$-decomposition if $C_i \cap C_j = \emptyset$ for $i \neq j$.

A $C^*$-decomposition partitions not only the set of reactions but also the set of complexes. The primary example of a $C^*$-decomposition are the linkage classes. Linkage classes, in fact, essentially determine the structure of a $C^*$-decomposition. We first present this Structure Theorem for a $C^*$-decomposition:

**Theorem 5. (Structure Theorem for $C^*$-decomposition).** Let $L_1, \ldots, L_l$ be the linkage classes of a network $N$. A decomposition $N = N_1 + N_2 + \ldots + N_k$ is a $C^*$-decomposition if and only if each $N_i$ is the union of linkage classes and each linkage class is contained in only one $N_i$. In other words, the linkage class decomposition is a refinement of $N$.

**Proof.** Clearly, if the linkage classes form a refinement of $N$, then $N$ is a $C^*$-decomposition. To see the converse, let $N_i = (S_i, C_i, R_i)$ and $L_j = (S_{L_j}, C_{L_j}, R_{L_j})$ where $R_i$ is the union (taken over $j$) of $(R_i \cap R_{L_j})$. We only need to show that each non-empty intersection is equal to $R_{L_j}$, (i.e., $R_{L_j} = R_i \cap R_{L_j}$) to imply that each linkage class is contained in only one $N_i$. If the linkage class $L_j$ has only one reaction then $R_i \cap R_{L_j} = R_{L_j}$. If the linkage class $L_j$ has at least two reactions, then there is an adjacent reaction to each reaction, whose reactant complex or product complex is common with the first reaction. If this adjacent reaction belongs to a different subnetwork, then there exists a complex which is common to two different subnetworks. This would contradict that $N$ partitions the set of complexes. Hence, all reactions of the linkage class lie in the intersection with $R_i$. □

**Corollary 4.** For a $C^*$-decomposition $N = N_1 + N_2 + \ldots + N_k$, $k \leq l$.

**Proof.** If $N$ is decomposed according to linkage classes, then $N_i = L_j$. Thus, $k = l$. If each $N_i$ is the union of linkage classes, then the number of subnetworks is less than the number of linkage classes. Hence, $k < l$. □

We also obtain a new characterization of the ILC property:
Corollary 5. A network has independent linkage classes if and only if every $C$-decomposition is independent.

Proof. If a network has independent linkage classes then the stoichiometric subspace $S$ is the direct sum of the stoichiometric subspaces of the linkage classes. Grouping the summands according to the unions of the linkage classes for the subnetworks of a $C$-decomposition provides $S$ as the direct sum of the subnetworks. Hence, every $C$-decomposition of a network is also independent. For the converse, since every $C$-decomposition of a network is independent and the linkage class decomposition is also a $C$-decomposition, it follows that the linkage classes are independent. □

If a network has dependent linkage class, it may fail to have an independent $C$-decomposition, as the following example shows:

Example 8. Consider the CRN with reactions $X_1 \rightarrow 2X_1 + X_2$ and $X_2 \rightarrow 2X_2 + X_1$, it has $\delta = 1$. The only non-trivial decomposition is the linkage class decomposition, where the deficiency of the two linkage classes is 0. Clearly, the linkage class decomposition is dependent. In particular, it has no independent $C$-decomposition.

We now derive the Structure Theorem for $C^*$-decomposition and apply this to the incidence-independence of such decomposition.

Theorem 6. (Structure Theorem for $C^*$-decomposition). Let $N_1 + N_2 + ... + N_k$ be a $C^*$-decomposition and $L_0$ and $L_{0,i}$ be the linkage classes of $N$ and $N_i$ containing the zero complex (note $L_{0,i}$ is empty if $N_i$ does not contain the zero complex). Then

i) the $L_{0,i}$ form a $C^*$-decomposition of $L_0$

ii) the (non-empty) $N_i \setminus L_{0,i}$ form a $C$-decomposition of $N \setminus L_0$

Proof. To prove (i), we need to show that each non-zero complex of $L_0$ is contained in only one subnetwork $N_i$. If there is only one subnetwork $N_i$ containing the zero complex then we are done. If there are at least two subnetworks containing the zero complex then $L_0$ has at least two non-zero complexes connected to the zero complex. Otherwise, if there would only be one complex then $N_i$ is not a $C^*$-decomposition of $N$, a contradiction. Now, if one of these non-zero complexes belongs to different subnetworks, this would contradict that $N$ partitions the non-zero complexes. Hence, all the non-zero complexes of $L_0$ is contained in only one $N_i$ and $L_{0,1} + L_{0,2} + ... + L_{0,j} = L_0$ for $j \leq k$. 

23
To prove (ii), it suffices to show that the intersection of the set of complexes in $N_i \setminus L_0$ is empty. The set of complexes in $N \setminus L_0$ are all non-zero and $N \setminus L_0 = (N_1 + ... + N_k) \setminus L_0$. From (i), we have $L_{0,1} + L_{0,2} + ... + L_{0,j} = L_0$ for $j \leq k$. Thus, $N \setminus L_0 = N_1 \setminus L_{0,1} + ... + N_k \setminus L_{0,k}$ where $L_{0,k}$ is empty if $N_k$ does not contain the zero complex. Since $N_i$ is a $C^*$-decomposition of $N$, the intersection of the set of complexes in $N_i \setminus L_{0,i}$ is empty.

Incidence-independence is given by Equation (1) in Section 4. For $C^*$-decompositions, this equation can be transformed into a more convenient “Common Complex Criterion” as follows:

**Proposition 6. (Common Complex Criterion).** Let $k(0)$ be the number of subnetworks $N_i$ in a network $N$ containing the zero complex. A $C^*$-decomposition of $N$ is incidence-independent if and only if

$$\sum l_i - l = \begin{cases} 0, & \text{whenever } k(0) = 0 \\ k(0) - 1, & \text{whenever } k(0) > 0 \end{cases}$$

**Proof.** We recall Equation (1):

$$n - l = \sum (n_i - l_i)$$

where $n_i$ and $l_i$ are number of complexes and linkage classes in the subnetwork $N_i$. Obviously, if the network $N$ does not contain the zero complex then $k(0) = 0$. Since $N$ is a $C^*$-decomposition, $n = \sum n_i$. Thus, Equation (1) is reduced to $\sum l_i - l = 0$.

On the other hand, if $k(0)$ subnetworks contain the zero complex then, $n = n^* + 1$ where $n^*$ is the number of non-zero complexes in $N$. Also, $n_i = n^*_i + 1$ where $n^*_i$ is the number of non-zero complexes in $N_i$. Thus, Equation (1) becomes

$$n^* + 1 - l = \sum n^*_i + k(0) - \sum l_i.$$ 

Since $n^* = \sum n^*_i$, we obtain $\sum l_i - l = k(0) - 1$. □

Using the parameter $k(0)$, we obtain two sufficient conditions for incidence-independence of $C^*$-decomposition.

**Proposition 7.** Let $k(0)$ be the number of subnetworks of a $C^*$-decomposition containing the zero complex.
i) If \( k(0) = 0 \) or \( 1 \), then the decomposition is a \( C \)-decomposition, and hence incidence-independent.

ii) If \( k(0) = k \), i.e. all subnetworks contain the zero complex, then the \( C^* \)-decomposition is incidence-independent.

Proof. \( k(0) = 0 \) means that the network does not contain the zero complex, by definition, the decomposition is a \( C \)-decomposition. This is also the case when the zero complex is in only one subnetwork in a \( C^* \)-decomposition. The incidence-independence also follows from the Common Complex Criterion (CCC).

From the Structure Theorem of \( C^* \)-decomposition (Theorem 6), there are precisely \( k \) subnetwork linkage classes in \( L_0 \). Since the subnetworks \( N_i \setminus L_0 \) form a \( C \)-decomposition of \( N \setminus L_0 \), and there is no zero complex in \( N \setminus L_0 \), \( \sum (l_i - 1) - (l - 1) = 0 \) or \( \sum l_i - l = k - 1 \), which is the CCC condition for incidence-independence of the \( C^* \)-decomposition of the whole network. \( \square \)

6.2. The new \( S \)-system CRN \( N^* \)

As mentioned in Section 3, the species set partition of the NRP and STR full models \( N \) generates embedded networks \( N_i^* \). But these \( N_i^* \)'s does not partition the reaction set in \( N \). This led us in constructing a new \( S \)-system \( N^* \).

Definition 18. Let \( N_i^* = (S_i, C|_{S_i}, R|_{S_i}) \), \( i = 1, ..., k \), be the embedded networks induced by the species and reaction set partitions. The \( * \)-disjoint union \( N^* \) of the embedded networks is the \( S \)-system CRN given by \( (S \cup C|_{S} \cup R|_{S}) \).

The adjective "\( * \)-disjoint" stems from the fact that if the zero complex is present in the network, the union of all complexes may not be disjoint. However if \( C^* \) denotes the set of non-zero complexes, then their union is disjoint. The partitions of the species set and the reaction set always hold.

The basic result on \( N^* \) is:

Proposition 8. The embedded networks \( N_i^* \) of \( N \) constitute an independent decomposition of \( N^* \).

Proof. The embedded networks \( N_i^* \) are subnetworks of \( N^* \) induced by the partition of the reaction set. Since their ranks sum up to the rank of \( N^* \), the decomposition is independent. \( \square \)

It follows from Proposition 3.3 in [9] that the deficiency of \( N^* \) is less than or equal to the sum of the embedded network deficiencies. It also follows from
Table 5: NRP and STR Network numbers for $N^*$ and the embedded networks $N_i^*$

| Network numbers | NRP $N^*$ | STR $N^*$ |
|-----------------|-----------|-----------|
| $m$             | 40 8 17 15 | 37 7 14 16 |
| $m_{rev}$       | 0 0 0 0    | 2 0 1 1    |
| $n$             | 95 19 36 42| 90 19 29 44|
| $n_r$           | 58 10 20 28| 57 12 17 29|
| $r$             | 80 16 34 30| 74 14 28 32|
| $l$             | 16 3 3 12  | 18 5 2 13  |
| $sl$            | 95 19 36 42| 90 19 28 43|
| $t$             | 37 9 16 14 | 33 7 12 15 |
| $s$             | 40 8 17 15 | 37 7 14 16 |
| $q$             | 40 8 17 15 | 37 7 14 16 |
| $\delta$        | 39 8 16 15 | 35 7 13 15 |
| $\delta_p$      | 18 2 3 13  | 20 5 3 13  |

Feinberg’s Decomposition Theorem, that the positive equilibria of $N^*$ consists of the elements in the intersection of the positive equilibria sets of the embedded networks.

**Example 9.** The embedded networks $N_1^*$, $N_2^*$ and $N_3^*$ of the NRP and STR full model constitute an independent decomposition of $N^*$ which can be verified from Table 5. Moreover, these embedded networks form a $G^*$-decomposition of $N^*$ with $k(0) = k = 3$. Hence, it follows from Proposition 7 that $N^*$ is incidence-independent. Thus, the $N_i^*$’s of the NRP and STR full model is a bi-independent $G^*$-decomposition of $N^*$. (This is summarized in Table 7)

7. The species subsets-induced decomposition of $N$ and modularity

In this section, we consider $S$-systems, which, like the Mtb NRP and STR models are derived from connected digraphs, with no independent variables and with a vertex partition defining subdigraphs.

7.1. The species subsets-induced decomposition of $N$

An $S$-system species set partition determines not only a set of embedded networks, but also a coarsening of the $S$-system’s species decomposition. Returning to the full network $N$, the reaction set partition also induces a decomposition into $k$ subnetworks $N_1, ..., N_k$. These subnetworks are not CRNs of $S$-systems since
they have species which are neither dependent nor independent variables. However, they are quite similar, and, in particular $N_i$ has rank $m$. Hence, we have the next basic fact:

**Proposition 9.** The subnetworks $N_i$ constitute an independent decomposition of $N$. It is incidence-independent if the species decomposition is incidence-independent.

**Proof.** The proof for the independence of $N_i$ is similar to Proposition [8]. Since the subnetworks $N_i$ is a coarsening of the species decomposition, the second claim follows from Proposition [3]. □

**Remark 5.** The converse of the incidence-independent claim does not follow. It has been shown in Example [4] that the species decomposition of NRP $N$ is incidence-independent but Table 6 shows that the network decomposition of $N = N_1 + N_2 + N_3$ is incidence-independent (i.e., $n - l = \sum n_i - l_i$).

We first relate the decomposition of $N$ with that of $N^*$. The relation of $N$ and $N^*$ is based on the observation that for each $i$ there is a digraph homomorphism from $N_i$ to $N_i^*$. A digraph homomorphism maps the vertices while preserving adjacency. For a complex in $N_i$ its image is its “projection” to $N_i^*$, i.e. the terms with common species are left out.

**Proposition 10.** The map $\phi : N_i \rightarrow N_i^*$ is a digraph homomorphism. Each such map is surjective, and if injective, is an isomorphism of the subdigraphs.

**Proof.** The construction of $i^{th}$ embedded network involves removing species not in the species set $S_i$ from the complexes of $N_i$. Hence there is a surjective map $\phi : R_i \rightarrow R_i^*$. Since any complex is a reactant or product of a reaction, one obtains a map $\phi : C_i \rightarrow C_i^*$ too. This map is clearly a digraph homomorphism. □

If the decomposition of $N$ is a $C^*$-decomposition, then the subdigraph homomorphism extend to a digraph homomorphism from $N$ to $N^*$. If, in addition, all the subdigraph maps are injective, the map $N$ to $N^*$ is a digraph isomorphism.

**Example 10.** The subnetworks $N_i$ is surjective digraph homomorphic to the embedded networks $N_i^*$ in both the NRP and STR models. Moreover, the subnetwork $N_2 (N_i)$ is injective to the embedded network $N_2^* (N_i^*)$ in NRP (STR).
Table 6: NRP and STR Network numbers for $\mathcal{N}$ and its subnetworks $\mathcal{N}_i$

| Network numbers | NRP $\mathcal{N}$ | $\mathcal{N}_1$ | $\mathcal{N}_2$ | $\mathcal{N}_3$ | STR $\mathcal{N}$ | $\mathcal{N}_1$ | $\mathcal{N}_2$ | $\mathcal{N}_3$ |
|----------------|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| $m$            | 40              | 11             | 18             | 17             | 37             | 9              | 19             | 19             |
| $m_{rev}$      | 0               | 0              | 0              | 0              | 0              | 0              | 0              | 0              |
| $n$            | 98              | 21             | 36             | 43             | 95             | 19             | 33             | 46             |
| $n_r$          | 60              | 12             | 20             | 28             | 60             | 12             | 19             | 29             |
| $r$            | 80              | 16             | 34             | 30             | 74             | 14             | 28             | 32             |
| $l$            | 19              | 5              | 3              | 13             | 22             | 5              | 6              | 14             |
| $sl$           | 98              | 21             | 36             | 43             | 94             | 19             | 32             | 46             |
| $t$            | 38              | 9              | 16             | 15             | 35             | 7              | 13             | 17             |
| $s$            | 40              | 8              | 17             | 15             | 37             | 7              | 14             | 16             |
| $q$            | 40              | 11             | 18             | 17             | 37             | 9              | 19             | 19             |
| $\delta$       | 39              | 8              | 16             | 15             | 36             | 7              | 13             | 16             |
| $\delta_p$    | 20              | 1              | 2              | 11             | 23             | 3              | 0              | 10             |

We now compare the network numbers of $\mathcal{N}$ and $\mathcal{N}^*$ for both the NRP and STR models. Some of the network numbers of $\mathcal{N}^*$ and $\mathcal{N}$ in Tables 5 and 6, respectively, are equal. In particular, their number of reactions ($r$) are equal however, their set of reactions are not the same. The reactions in $\mathcal{N}^*$ that differ from $\mathcal{N}$ is indicated by a reaction $R_i^*$ in the Supplementary Information. Further, there are 3 more complexes in $\mathcal{N}$ than in $\mathcal{N}^*$. This is due to the construction of $\mathcal{N}^*$ which is from the union of the embedded networks $\mathcal{N}_i^*$ that involves removal of species not in the species set $S_i$ from the complexes of $\mathcal{N}_i^*$ which leads to coincidence of complexes. Thus, this reduced number of complexes in $\mathcal{N}^*$ also reduces the other network numbers: $n_r$, $l$, $sl$ and $t$ in $\mathcal{N}^*$.

Meanwhile, the network numbers of $\mathcal{N}_i^*$ and $\mathcal{N}_i$ have also similarities and differences. Their $r$, $s$ and $\delta$ are equal. There are also more species in $\mathcal{N}_i$ than in $\mathcal{N}_i^*$ because of an addition of “foreign species” (those not in the subsystem).

In terms of the network properties, both $\mathcal{N}^*$ and $\mathcal{N}$ with their subnetworks $\mathcal{N}_i^*$ and $\mathcal{N}_i$, respectively, all have the same network properties (this properties are similar to those presented in Tables 2 and 3).

The embedded networks (which corresponds to the gene subsystems) $\mathcal{N}_i^*$ in $\mathcal{N}^*$ form a $\mathcal{C}^*$-decomposition (Table 7). This means that their only interactions are “outside” the system as represented by the zero complex. The subnetworks $\mathcal{N}_i$ in $\mathcal{N}$ on the other hand, have reactions where foreign species are involved and it even have complexes in common other than the zero complex. In fact, the
Table 7: Description of the network decomposition of $\mathcal{N}$ and $\mathcal{N}^*$

| Network decomposition | Independent | Incidence-independent | $\mathcal{C}^*$-decomposition |
|-----------------------|-------------|-----------------------|-------------------------------|
| NRP                   | YES         | YES                   | YES                           |
| $\mathcal{N}^* = \mathcal{N}_1^* + \mathcal{N}_2^* + \mathcal{N}_3^*$ |
| $\mathcal{N} = \mathcal{N}_1 + \mathcal{N}_2 + \mathcal{N}_3$ | YES         | YES                   | YES                           |
| STR                   | YES         | YES                   | NO                            |
| $\mathcal{N} = \mathcal{N}_1 + \mathcal{N}_2 + \mathcal{N}_3$ | YES         | YES                   | NO                            |
| $\mathcal{N}^* = \mathcal{N}_1^* + \mathcal{N}_2^* + \mathcal{N}_3^*$ | YES         | YES                   | YES                           |

non-zero complex “Y10” is common in the STR $\mathcal{N}_2$ and $\mathcal{N}_3$ which makes the STR $\mathcal{N} = \mathcal{N}_1 + \mathcal{N}_2 + \mathcal{N}_3$ not a $\mathcal{C}^*$-decomposition (Table 7).

7.2. Modularity of digraph divisions and modularity of CRN decompositions

Biological systems often display an organization into functional modules and hence, it is important that models of such systems capture characteristics of the modular structure. In this Section, we discuss the modularity concepts available for digraph divisions, but provide evidence that these need to be modified or extended for CRN decompositions.

Modularity was initially introduced by Newman and Girvan for the case of undirected networks while Arenas proposed an extension of this for directed networks [13]. Their extension is based on the observation that the existence of a directed edge $(i, j)$ between nodes $i$ and $j$, depends on the out-degree and in-degree of nodes $i$ and $j$ respectively. The modularity for directed network, denoted by $Q$, is expressed as:

$$Q = \frac{1}{m} \sum_{i,j} [A_{ij} - \frac{k_{out}^i k_{in}^j}{m}] \delta(c_i, c_j)$$

where $m$ is the total number of arcs in the network, $A_{ij}$ is the number of arcs from $i$ to $j$, $k_{out}^i$ and $k_{in}^j$ are the outdegree and indegree of the nodes $i$ and $j$, respectively and $\delta(c_i, c_j) = 1$ (i.e., if nodes $i$ and $j$ belong on the same module) and 0 otherwise.

In terms of the structure or graph of a network, modularity is designed to measure the strength of division of a network into modules (clusters or communities). Good divisions, which have high modularity values, are those with dense edge connections between the vertices within a module but sparse connections between vertices in different modules [13].

A systems biologist describes modules from a graph-theoretical point of view as a group of nodes that are more strongly intraconnected than interconnected
while a geneticist might consider a set of co-expressed or co-regulated genes a module [14].

When we compute the modularities $Q$ of the divisions of the digraphs in Figure 2 and the reaction graph of the CRN decompositions of $\mathcal{N}$ (which is also a digraph) for NRP and STR, we obtain the following surprising results:

| Mtb S-system | $Q$ (digraph)  | $Q$ (reaction graph) |
|--------------|----------------|----------------------|
| NRP          | 0.4405         | 0.3217               |
| STR          | 0.4472         | 0.1989               |

Paradoxically, the digraph model indicates that STR is more modular than NRP, while the same measure applied to the reaction graph of the CRN indicates the opposite. Our conclusion from this computation is that because a chemical reaction network has a richer structure than just a digraph, one needs to modify or expand the concept of modularity for a CRN to include aspects of its stoichiometric structure. In the following, we introduce an initial concept which is admittedly specific for $S$-system CRNs but might lead in the right direction.

Now, if two subnetworks (i.e. subdigraphs) are in different connected components, then they are “physically” isolated, and hence, the question of modularity is trivial. We hence assume in the following that the subdigraphs in question are in the same connected component.

Any arc in the digraph (biochemical map) whose source and target vertices lie in different subdigraphs lead to the occurrence of common species between the two subnetworks of $\mathcal{N}$ in the $S$-system CRN’s. Since by assumption, the digraph under consideration is connected, for each subdigraph, there is at least one such arc and hence at least one common species with another subnetwork. Based on this, we introduce the concept of species coupling level of a subnetwork:

**Definition 19.** The *species coupling level* $c_S(N')$ of a subnetwork $N'$ in $\mathcal{N}$ is the ratio of the number of occurrences of common species in the reactant and product complexes of the input and output reactions, respectively and the number of occurrences of non-common species in the reactant complexes of the input reactions.
With respect to the digraph of the S-system, the numerator counts the number of arcs between vertices from different subdigraphs coming into the subdigraph while the denominator counts the number of arcs between vertices within the subdigraph.

**Remark 6.** The species coupling level of a network is equal to the sum of the species coupling levels of its subnetworks.

Table 9 compares the modularity of the digraphs with the species coupling levels for NRP and STR.

Table 9: Modularity $Q$ of the digraphs and the species coupling levels $\sum c_s(N_i)$ for NRP and STR

| Mtb S-system | $Q$ (digraph) | $\sum c_s(N_i)$ |
|--------------|---------------|-----------------|
| NRP          | 0.4405        | 0.6236          |
| STR          | 0.4472        | 0.5289          |

Since modularity is inversely related to species coupling, we see that the values of the latter for NRP and STR are qualitatively consistent with the modularity values for the corresponding digraphs. In our view, this indicates that stoichiometric level information is useful for modularity considerations.

The NRP and STR decompositions contain only mono-species common complexes. We counted the single non-zero one in the species-level coupling calculation and considered the common zero complex as “outside of the system” studied. We would view the occurrence of common multi-species complexes in other examples as further indicators on the stoichiometric level of lower modularity. We hope to explore these and related concepts in a more general context of developing an appropriate modularity concept for chemical reaction network decompositions.

8. **Conclusions and outlook**

Motivated by Magombedze and Mulders [15] approach in representing and analyzing the gene regulatory based system model of *Mtb* in modular form, where the system partitions the vertices into subsets that form corresponding subsystems, we studied these subsystems as embedded networks. In the embedded CRN representation of S-system, it is defined as a species subset and a reaction subset induced by the set of digraph vertices of the subsystem.

In this study, we have
1. corrected Proposition 10 of [1] and provided a new proof using decomposition theory.

2. presented the discordance of $S$-system CRN with at least two dependent species.

3. developed the concept of $C$- and $C^*$-decompositions including their structure theorems in terms of linkage classes and defined a new $S$-system CRN $N^*$. The union of the embedded networks $N_i^*$ of $N$ construct the new $S$-system CRN $N^*$. Note that these $N_i^*$'s do not partition the reaction set in $N$ but constitute an independent $C^*$-decomposition of $N^*$.

4. shown that the species subsets-induced decomposition of $N = \sum N_i$ induces surjective digraph homomorphisms between $N_i$ and $N_i^*$.

5. illustrated that the modularity of the decomposition of the $S$-system $N$ is not consistent with the modularity of the original digraph model of the gene regulatory system. In order for the modularity concept from digraph to capture the stoichiometric structure of CRN, we have introduced the concept of species coupling level for the CRN decompositions.

Biological systems are complex networks that exhibit the orchestrated interplay of a large array of components [12]. It is common practice for modellers to decompose the complex system into subsystems. It is believed that studying the dynamics and functionality of the subsystems would facilitate understanding of the system as a whole. Henceforth, discovering and analyzing such subsystems are crucial in gaining better understanding of the complex systems [12]. There are still a lot to explore and study regarding the topological/network properties with respect to decomposition/separability of subsystems.

Acknowledgments. HFF acknowledges the support of the Commission on Higher Education (CHED), Philippines for the CHED-SEGS Scholarship Grant. ARL held research fellowship from De La Salle University and would like to acknowledge the support of De La Salle Universitys Research Coordination Office.
References

[1] C.P.P. Arceo, E.C. Jose, A. Marin-Sanguino, E.R. Mendoza, Chemical reaction network approaches to biochemical systems theory, Math. Biosci. 269 (2015) 135152.

[2] C.P.P. Arceo, E.C. Jose, A.R. Lao, E.R. Mendoza, Reaction networks and kinetics of biochemical systems, Math. Biosci. 283 (2017) 13-29.

[3] C.P.P. Arceo, E.C. Jose, A.R. Lao, E.R. Mendoza, Reactant subspaces and kinetics of biochemical systems, J. Math. Chem. 283 (2017) 13-29.

[4] B. Boros, On the positive steady states of Deficiency-One Mass Action Systems, Eotvos Lorand University (2013) Ph.D. thesis

[5] M. Domijan, M. Kirkilionis, Graph theory and qualitative analysis of reaction networks, Networks Heterog. Media. 3 (2008) 295-322.

[6] M. Feinberg, Chemical reaction network structure and the stability of complex isothermal reactors: I. The deficiency zero and deficiency one theorems, Chem. Eng. Sci. 42 (1987) 2229-2268.

[7] M. Feinberg, F.J.M. Horn, Chemical mechanism structure and the coincidence of the stoichiometric and kinetic subspaces, Arch. Rational Mech. Anal. 66 (1977) 8397.

[8] E. Feliu, C. Wiuf, Preclusion of switch behaviour in networks with mass action kinetics, Appl. Math. Comput. 219 (2012) 14491467.

[9] N.T. Fortun, A.R. Lao, L.F. Razon, E.R. Mendoza, A Deficiency Zero Theorem for a class of power law kinetic systems with independent decompositions, 2018 (submitted).

[10] B. Joshi, A. Shiu, Simplifying the Jacobian Criterion for precluding multistationarity in chemical reaction networks, SIAM J. Appl. Math. 72 (2012) 857876.

[11] B. Joshi, A. Shiu, Atoms of multistationarity in chemical reaction networks, J. Math. Chem. 51 (2013) 153-178.
[12] P.M.S. Kim, Understanding Subsystems in Biology through Dimensionality Reduction, Graph Partitioning and Analytical Modeling, (2003). http://hdl.handle.net/1721.1/7099

[13] W. Li, D. Schuurmans, Modular Community Detection in Networks, IJCAI Proceedings-International Jt. Conf. Arti. Intell. (2011) 1366-1371.

[14] D.M. Lorenz, A. Jeng, M.W. Deem, The emergence of modularity in biological systems, Phys. Life Rev. 8 (2011) 129-160.

[15] G. Magombedze, N. Mulder, Understanding TB latency using computational and dynamic modelling procedures, Infect. Genet. Evol. 13 (2013) 267-283.

[16] G. Shinar, M. Feinberg, Concordant chemical reaction networks, Math. Biosci. 240 (2012) 92-113.

[17] E.O. Voit, Computational Analysis of Biochemical Systems, Cambridge University Press, 2000.
Appendix A. Fundamentals of chemical reaction networks and kinetic systems

We recall the necessary concepts of chemical reaction networks and the mathematical notation used throughout the paper adopted from the papers [1], [6] and [9].

We begin with the definition of a chemical reaction network.

**Definition 20.** A chemical reaction network is a triple $N = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ of three non-empty finite sets:

1. A set species $\mathcal{S}$,
2. A set $\mathcal{C}$ of complexes, which are non-negative integer linear combinations of the species, and
3. A set $\mathcal{R} \subseteq \mathcal{C} \times \mathcal{C}$ of reactions such that
   - $(i, i) \notin \mathcal{R}$ for all $i \in \mathcal{C}$, and
   - For each $i \in \mathcal{C}$, there exists a $j \in \mathcal{C}$ such that $(i, j) \in \mathcal{R}$ or $(j, i) \in \mathcal{R}$.

We denote with $m$ the number of species, $n$ the number of complexes and $r$ the number of reactions in a CRN.

**Definition 21.** A complex is called monospecies if it consists of only one species, i.e., of the form $kX_i$, $k$ a non-negative integer and $X_i$ a species. It is called monomolecular if $k = 1$, and is identified with the zero complex for $k = 0$.

A zero complex represents the “outside” of the system studied, from which chemicals can flow into the system at a constant rate and to which they can flow out at a linear rate (proportional to the concentration of the species). In biological systems, the “outside” also stands for the degradation of a species.

A chemical reaction network $(\mathcal{S}, \mathcal{C}, \mathcal{R})$ gives rise to a digraph with complexes as vertices and reactions as arcs. However, the digraph determines the triple uniquely only if an additional property is considered in the definition: $\mathcal{S} = \bigcup \text{supp } i$ for $i \in \mathcal{C}$, i.e., each species appears in at least one complex. With this additional property, a CRN can be equivalently defined as follows.

**Definition 22.** A chemical reaction network is a digraph $(\mathcal{C}, \mathcal{R})$ where each vertex has positive degree and stoichiometry, i.e., there is a finite set $\mathcal{S}$ (whose elements are called species) such that $\mathcal{C}$ is a subset of $\mathbb{R}_+^\mathcal{S}$. Each vertex is called a complex and its coordinates in $\mathbb{R}_+^\mathcal{S}$ are called stoichiometric coefficients. The arcs are called reactions.
Two useful maps are associated with each reaction:

**Definition 23.** The reactant map \( \rho : \mathcal{R} \rightarrow \mathcal{C} \) maps a reaction to its reactant complex while the product map \( \pi : \mathcal{R} \rightarrow \mathcal{C} \) maps it to its product complex. We denote \(|\rho(\mathcal{R})| \) with \( n_r \), i.e., the number of reactant complexes.

Connectivity concepts in Digraph Theory apply to CRNs, but have slightly differing names. A connected component is traditionally called a **linkage class**, denoted by \( \mathcal{L} \), in CRNT. A subset of a linkage class where any two elements are connected by a directed path in each direction is known as a **strong linkage class**. If there is no reaction from a complex in the strong linkage class to a complex outside the same strong linkage class, then we have a **terminal strong linkage class**. We denote the number of linkage classes with \( l \), that of the strong linkage classes with \( sl \) and that of terminal strong linkage classes with \( t \). Clearly, \( sl \geq t \geq l \).

Many features of CRNs can be examined by working in terms of finite dimensional spaces \( \mathbb{R}^{\mathcal{S}}, \mathbb{R}^{\mathcal{C}}, \mathbb{R}^{\mathcal{R}} \), which are referred to as species space, complex space and reaction space, respectively. We can view a complex \( j \in \mathcal{C} \) as a vector in \( \mathbb{R}^{\mathcal{C}} \) (called **complex vector**) by writing \( j = \sum_{s \in \mathcal{S}} j_s s \), where \( j_s \) is the stoichiometric coefficient of species \( s \).

**Definition 24.** The reaction vectors of a CRN \((\mathcal{S}, \mathcal{C}, \mathcal{R})\) are the members of the set \( \{ j - i \in \mathbb{R}^{\mathcal{S}} | (i, j) \in \mathcal{R} \} \). The **stoichiometric subspace** \( S \) of the CRN is the linear subspace of \( \mathbb{R}^{\mathcal{S}} \) defined by

\[
S : \text{span}\{ j - i \in \mathbb{R}^{\mathcal{S}} | (i, j) \in \mathcal{R} \}.
\]

The **rank** of the CRN, \( s \), is defined as \( s = \text{dim}S \).

**Definition 25.** The incidence map \( I_a : \mathbb{R}^{\mathcal{R}} \rightarrow \mathbb{R}^{\mathcal{C}} \) is defined as follows. For \( f : \mathcal{R} \rightarrow \mathbb{R} \), then \( I_a(f)(v) = -f(a) \) and \( f(a) \) if \( v = \rho(a) \) and \( v = \pi(a) \), respectively, and are 0 otherwise.

Equivalently, it maps the basis vector \( \omega_a \) to \( \omega_{v'} - \omega_v \) if \( a : v \rightarrow v' \).

It is clearly a linear map, and its matrix representation (with respect to the standard bases \( \omega_a, \omega_v \)) is called the **incidence matrix**, which can be described as

\[
(I_a)_{i,j} = \begin{cases} 
-1 & \text{if } \rho(a_j) = v_i, \\
1 & \text{if } \pi(a_j) = v_i, \\
0 & \text{otherwise.}
\end{cases}
\]
Note that in most digraph theory books, the incidence matrix is set as $-I_a$.

An important result of digraph theory regarding the incidence matrix is the following:

**Proposition 11.** Let $I$ be the incidence matrix of the directed graph $D = (V, E)$. Then \( \text{rank } I = n - l \), where $l$ is the number of connected components of $D$.

A non-negative integer, called the deficiency, can be associated to each CRN. This number has been the center of many studies in CRNT due to its relevance in the dynamic behavior of the system.

**Definition 26.** The **deficiency** of a CRN is the integer $\delta = n - l - s$.

**Definition 27.** The **reactant subspace** $R$ is the linear space in $\mathbb{R}^S$ generated by the reactant complexes. Its dimension, $\text{dim } R$ denoted by $q$, is called the **reactant rank** of the network. Meanwhile, the **reactant deficiency** $\delta_p$ is the difference between the number of reactant complexes and the reactant rank, i.e., $\delta_p = n_r - q$.

We now introduce the fundamentals of chemical kinetic systems. We begin with the general definitions of kinetics from Feliu and Wiuf [8]:

**Definition 28.** A **kinetics** for a CRN $(\mathcal{S}, C, R)$ is an assignment of a rate function $K_j : \Omega_K \rightarrow \mathbb{R}_{\geq}$ to each reaction $r_j \in R$, where $\Omega_K$ is a set such that $\mathbb{R}_{\geq} \subseteq \Omega_K \subseteq \mathbb{R}_{\geq}$ whenever $c, d \in \Omega_K$, and

\[
K_j(c) \geq 0, \quad \forall c \in \Omega_K.
\]

A kinetics for a network $\mathcal{N}$ is denoted by $K = (K_1, K_2, ..., K_r) : \Omega_K \rightarrow \mathbb{R}_{\geq}^R$. The pair $(\mathcal{N}, K)$ is called the **chemical kinetic system** (CKS).

In the definition, $c \wedge d$ is the bivector of $c$ and $d$ in the exterior algebra of $\mathbb{R}^S$. We add the definition relevant to our context:

**Definition 29.** A chemical kinetics is a kinetics $K$ satisfying the positivity condition: for each reaction $j : y \rightarrow y', K_j(c) > 0$ iff $\text{supp } y \subset \text{supp } c$.

Once a kinetics is associated with a CRN, we can determine the rate at which the concentration of each species evolves at composition $c$.

Power-law kinetics is defined by an $r \times m$ matrix $F = [F_{ij}]$, called the **kinetic order matrix**, and vector $k \in \mathbb{R}^m$, called the **rate vector**. In power-law formalism, the kinetic orders of the species concentrations are real numbers.

**Definition 30.** A kinetics $K : \mathbb{R}_{\geq}^\mathcal{S} \rightarrow \mathbb{R}^\mathcal{S}$ is a **power-law kinetics** (PLK) if

\[
K_j(x) = k_i x^{F_i}, \quad \forall i = 1, ..., r
\]

with $k_i \in \mathbb{R}_{\geq}$ and $F_{ij} \in \mathbb{R}$.
Appendix B. Network decomposition

The following diagram illustrates several decompositions introduced in this paper showing its relationship to Feinberg’s decomposition theory.

Figure B.4: Relationship and properties of Feinberg’s decomposition theory (in solid-lined boxes) to the new terms (in dash-lined boxes) that emerged from this paper.