

A Topological Approach to Chemical Organizations

Gil Benkö^{*,**}

University of Leipzig

Florian Centler[†]

Friedrich Schiller University Jena

Peter Dittrich[†]

Friedrich Schiller University Jena

Christoph Flamm^{**,‡}

University of Leipzig

University of Vienna

Bärbel M. R. Stadler[§]

Max Planck Institute for Mathematics
in the Sciences

Peter F. Stadler^{**,‡,¶}

University of Leipzig

University of Vienna

The Santa Fe Institute

Abstract Large chemical reaction networks often exhibit distinctive features that can be interpreted as higher-level structures. Prime examples are metabolic pathways in a biochemical context. We review mathematical approaches that exploit the stoichiometric structure, which can be seen as a particular directed hypergraph, to derive an algebraic picture of chemical organizations. We then give an alternative interpretation in terms of set-valued set functions that encapsulate the production rules of the individual reactions. From the mathematical point of view, these functions define generalized topological spaces on the set of chemical species. We show that organization-theoretic concepts also appear in a natural way in the topological language. This abstract representation in turn suggests the exploration of the chemical meaning of well-established topological concepts. As an example, we consider connectedness in some detail.

Keywords

Chemical reaction networks, hypergraphs, stoichiometric matrix, flux analysis, generalized topology, closure functions, chemical organization theory, connectedness

1 Introduction

A distinguishing feature of chemistry is that the changes of molecules upon interaction are not limited to quantitative physical properties such as free energy, density, or concentrations, since molecular interactions do not only produce more of what is already there—rather, novel molecules can be generated. This is the principal difficulty for any theoretical treatment. Dynamical systems models of chemistry start from a fixed set of molecular species and trace the time dependence of their concentrations, in the

* Contact author.

** Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany. E-mail: gil@bioinf.uni-leipzig.de (G.B.); xtof@bioinf.uni-leipzig.de (C.F.); studla@bioinf.uni-leipzig.de (P.F.S.)

† Bio Systems Analysis Group, Jena Center for Bioinformatics and Department of Mathematics and Computer Science, Friedrich Schiller University Jena, D-07743 Jena, Germany. E-mail: dittrich@minet.uni-jena.de (P.D.)

‡ Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria. E-mail: xtof@tbi.univie.ac.at (C.F.); studla@tbi.univie.ac.at (P.F.S.)

§ Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, D-04103, Leipzig, Germany. E-mail: stadler@mis.mpg.de

¶ The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501.

same manner as population dynamics describe frequencies of genes or organisms. In this picture, molecules, genes, and biological species are reduced to mere indices of the concentration variables, while their properties are implicitly described by rate constants that must be determined outside these theories. On the other hand, one (or maybe the) interesting question in chemistry is *which* (novel) molecules arise in a given situation.

Walter Fontana and Leo Buss [19, 20] introduced the term *constructive dynamical system* for such a setting, in which (structural) innovation is a defining property of dynamics. In the context of a theory for biological organization, they identified and discussed in detail the lack of a theory for constructive dynamical systems. Their work emphasizes an algebraic point of view, in which the interactions (i.e., in the chemical context, the reactions) are viewed as algebraic operations. Fontana's *AlChem* [23, 20], for example, represents molecules as λ -calculus expressions. Reactions are defined by the operations of "application" of one λ -term to its reaction partner. The result is a new λ -term. It is worth noting in this context that chemical reactions can in turn be regarded as a model of computation, a possibility that is realized for example in the *chemical abstract machine* [6]. A wide variety of different computational paradigms have been used in artificial chemistry models, from strings and matrices to Turing machines and graphs [1, 2, 52, 14, 58, 57, 45, 5]. We refer to [13] for a review.

Large-scale chemical reaction networks arise in very different situations, from the metabolic networks of living cells [18] to the chemistry of planetary atmospheres [62] and combinatorial chemistry (see, e.g., [37]). Furthermore, artificial chemistry models at various levels of realism can be used to generate very large reaction networks, with the particular aim of determining generic properties [4]. This raises the question of analyzing large-scale structures in such networks.

The first approaches towards a systematic understanding of chemical reaction networks were based on an analysis of stoichiometry in mass-action reaction systems [17, 51, 8, 42, 34, 43]. Results such as the deficiency-zero theorem provide a link between network structure and reaction kinetics [16, 41, 40, 29]. Metabolic flux analysis [8, 18, 49, 35] developed analogous concepts, such as "flux modes" that describe the dominant reaction pathways within the network. From a mathematical point of view, the problem is to find the extremal rays of a cone that is determined by the stoichiometric matrix \mathbf{S} ; see below for details. The extremal flux vectors are closely related to the (directed) cycles of the network graph [27]. The theory of Petri nets [47] and oriented matroids [46] is closely related mathematically.

Particular classes of chemical reaction networks, namely those in which reactions are catalyzed by other members of the network, are of particular interest in the context of origin-of-life models [44, 38]. The emergence of autocatalytic sets is investigated in terms of a closure function that identifies self-sustaining sets where every reaction is catalyzed by a molecule in the set [39].

Based on Fontana's ideas, Peter Dittrich and Pietro Speroni di Fenizio recently presented a rigorous mathematical theory of chemical organizations based on notions of self-maintenance and closure [12] (for an application see, e.g., [7]). From a mathematical point of view, their work emphasizes lattice theoretic structures.

Concepts from point set topology appear to have been rarely applied to chemical reaction networks. Some of us argued, motivated by applications of topological concepts in evolutionary biology [21, 22, 10, 54, 53, 60], that a chemical system can be meaningfully described by a generalized closure function that determines what can be constructed from a given set of molecular species [55]. Such a closure function defines in a natural way a topological structure on a chemical reaction system. In this contribution we discuss the connections between chemical organizations and generalized topologies in some detail.

2 Chemical Reactions

Let X be a (finite or infinite) set of chemical species (molecular types). A *chemical reaction* on X is a pair of formal linear combinations of elements of X :

$$\rho : \sum_x \bar{\alpha}_{x,\rho} x \rightarrow \sum_x \alpha_{x,\rho} x, \quad \text{with } \alpha_{x,\rho}, \bar{\alpha}_{x,\rho} \geq 0. \quad (1)$$

For each reaction ρ we define its domain (the set of educts) and its image (the set of products):

$$\text{dom}\rho = \{x \in X \mid \bar{\alpha}_{x,\rho} > 0\}, \quad \text{img}\rho = \{x \in X \mid \alpha_{x,\rho} > 0\}. \quad (2)$$

We say that ρ is *catalyzed* if both $\bar{\alpha}_{x,\rho} > 0$ and $\alpha_{x,\rho} > 0$ for some $x \in X$. In this case, x is a *catalyst* of ρ . The set of catalysts of ρ is thus $\text{dom}\rho \cap \text{img}\rho$. Chemical reactions are often modeled as pairs of multisets (e.g., in [12]), corresponding to integer stoichiometric constants. However, there does not seem to be a technical or computational advantage associated with this restriction. Hence we allow arbitrary non-negative values of $\alpha_{x,\rho}$ and $\bar{\alpha}_{x,\rho}$.

A *chemical universe* is a pair (X, \mathfrak{R}) , where \mathfrak{R} is a set of chemical reactions on X . It can be represented by an arc-weighted bipartite directed graph $\Gamma(X, \mathfrak{R})$ with vertex set $X \cup \mathfrak{R}$, arcs $x \rightarrow \rho$ with weight $\bar{\alpha}_{x,\rho}$ if $\bar{\alpha}_{x,\rho} > 0$, and arcs $\rho \rightarrow x$ with weight $\alpha_{x,\rho} > 0$ if $\alpha_{x,\rho} > 0$. See Figure 1 for an example.

We will regard all reactions as irreversible. Reversible chemical reactions are thus represented as two different reactions ρ and $\bar{\rho}$ in \mathfrak{R} satisfying $\bar{\alpha}_{x,\bar{\rho}} = \alpha_{x,\rho}$ and $\alpha_{x,\bar{\rho}} = \bar{\alpha}_{x,\rho}$.

This framework is general enough to treat open multi-phase systems: Chemical species in different phases or compartments are treated as different molecular types. Transport between phases is represented by reactions of the form $\tau : x_i \rightarrow x_j$. A special type of transport reactions is influx into the system and efflux from the system, represented by $\varphi_x : \emptyset \rightarrow x$ and $\psi_x : x \rightarrow \emptyset$.

Note that we do not assume any internal structure of the chemical species x . Thus, our formalism does not incorporate conservation of mass or atom types. This is convenient, since it allows us to deal, for example, with *replicator systems* [36] of the form $I_k + I_l \rightarrow 2I_k + I_l$. To be chemically plausible, such a system must consume building material for the second copy of the template I_k and produce low-energy waste. One commonly assumes in this context that the building material is “buffered,” that is, supplied externally at a constant concentration, and that the waste is removed and does not influence the replication process. For our purposes, $I_k + I_l \rightarrow 2I_k + I_l$ is therefore a perfectly valid “chemical” reaction. The formalism developed here is therefore also applicable to artificial chemistry models.

Let us now consider an arbitrary subset $\mathcal{A} \subseteq X$, as exemplified by Figure 2. Clearly, a chemical reaction ρ can take place in a reaction mixture composed of the molecules in \mathcal{A} if and only if $\text{dom}\rho \subseteq \mathcal{A}$. The collection of all possible reactions in the universe (X, \mathfrak{R}) that can start from \mathcal{A} is thus given by

$$\mathfrak{R}_{\mathcal{A}} = \{\rho \in \mathfrak{R} \mid \text{dom}\rho \subseteq \mathcal{A}\}. \quad (3)$$

3 Stoichiometric Matrix and Flux Vectors

The first formal approach to analyzing the structure of complex reaction networks was probably Clarke’s stoichiometric network analysis [8]. As in related later developments such as metabolic flux

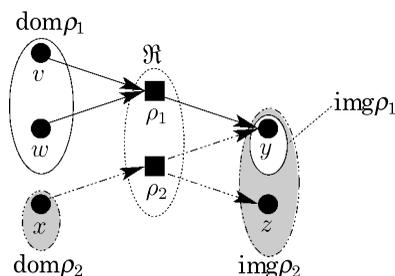


Figure 1. Illustration of the chemical universe (X, \mathfrak{R}) for the chemical reaction network composed of the two reactions $\rho_1 : v + w \rightarrow y$ and $\rho_2 : x \rightarrow y + z$. The chemical universe is the tuple of the set of species $X = \{v, w, x, y, z\}$ and the set of chemical reactions $\mathfrak{R} = \{\rho_1, \rho_2\}$.

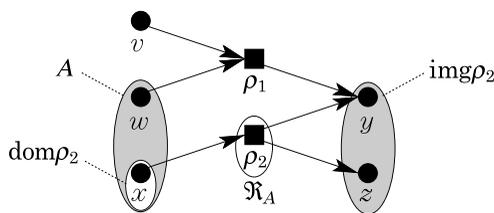


Figure 2. If only a subset A of all chemical species X is present in the reaction mixture, then only those reactions can take place whose educts $\text{dom}\rho_i$ are elements of A . This is true only for ρ_2 in the reaction network shown.

analysis (see, e.g., the book [18] and [48–50, 25] for recent developments in this area), its mathematical starting point is the stoichiometric matrix \mathbf{S} .

If X and \mathfrak{R} are finite, it makes sense to define the *stoichiometric matrix* \mathbf{S} of (X, \mathfrak{R}) with entries

$$s_{x,\rho} = \alpha_{x,\rho} - \bar{\alpha}_{x,\rho}. \quad (4)$$

Even if X and \mathfrak{R} are infinite, it is still meaningful to introduce a stoichiometric matrix \mathbf{S}_A on A in the same way for all $x \in A$ and $\rho \in \mathfrak{R}_A$ provided these two sets at least are finite.

We say that x is produced by the reaction ρ if $s_{x,\rho} > 0$, and that it is consumed by ρ is $s_{x,\rho} < 0$.

For our purposes it will also be convenient to work with a *modified stoichiometric matrix* $\hat{\mathbf{S}}_A$ of (X, \mathfrak{R}_A) that in addition includes efflux reactions for all $x \in X$. The entries of $\hat{\mathbf{S}}_A$ are $s_{x,\rho}$ for all $x \in X$ and all $\rho \in \mathfrak{R}_A$ except efflux reactions and additional entries $s_{x,\psi_x} = -1$ for the efflux reactions of all $x \in X$.

A *flux vector* J is a non-negative vector indexed by the reactions $\rho \in \mathfrak{R}$. In passing we note that several authors allow negative fluxes in the case of reversible reactions [35, 49]; instead, we represent reversible reactions as a pair $\rho, \bar{\rho}$. The entries $\mathbf{S}J$ are the production or consumption rates of chemical species, so that the kinetics of a chemical reaction can be always written in the form

$$\frac{d}{dt}[x] = (\mathbf{S}J)_x. \quad (5)$$

A flux vector J is called *stationary* if $\mathbf{S}J = \vec{0}$. Obviously, fixed points of the kinetic equations (5) imply stationary flux vectors. Note that in Equation 5 J depends on the current concentration vector changing in time, while \mathbf{S} is constant. See Figure 3 for an example.

We immediately observe the following relationships between \mathbf{S}_A and $\hat{\mathbf{S}}_A$: A flux vector J can be extended to a stationary flux vector w.r.t. $\hat{\mathbf{S}}_A$ if and only if $(\mathbf{S}_A J)_x \geq 0$ for all $x \in A$.

4 Chemical Organizations

In [12] a formal *theory of chemical organizations* is outlined, based on the following definitions:

1. A subset $A \subset X$ is “closed”¹ if for all reactions ρ with $\text{dom}\rho \subseteq A$ we have $\text{img}\rho \subseteq A$.
2. The “closure” $C(A)$ of a set A is the smallest “closed” set containing A . A similar closure function plays a crucial role in the theory of “catalytic reaction systems” by Mike Steel and Wim Hordijk [39].
3. A set A is *semi-self-maintaining* if every species x that is “consumed” (i.e., there is a reaction $\theta \in \mathfrak{R}_A$ such that $s_{x,\theta} < 0$) is also produced (i.e., there is a reaction $\rho \in \mathfrak{R}_A$ such that $s_{x,\rho} > 0$).

¹ We use quotes here to distinguish the definition of a closed set in the sense of the theory of chemical organizations from the topological notion of a closed set. As we shall see, these two definitions are related but not quite the same.

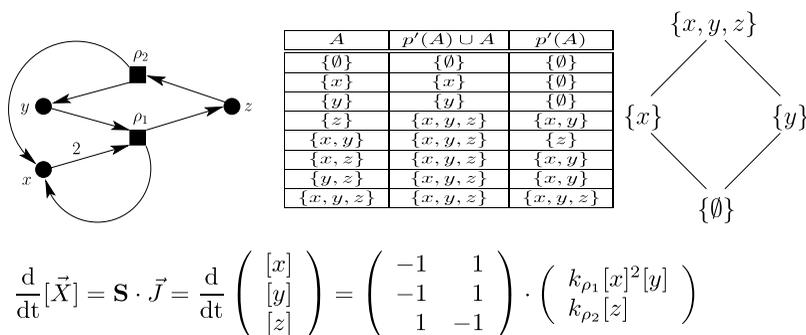


Figure 3. Left diagram: reaction network composed of the two reactions $\rho_1 : 2x + y \rightarrow z + x$ and $\rho_2 : z \rightarrow x + y$. Right diagram: the lattice of organizations for the reaction network. Note that catalysts (x in reaction ρ_1) vanish in the stoichiometric matrix \mathbf{S} , since the entry for a particular chemical species i is the difference $\bar{\alpha}_i - \alpha_i$ of the stoichiometric coefficients of the product and educt sides of the reaction equation. The table in the middle displays two closure functions p and p' applied to different sets A of chemical species.

- A set A is called a *semi-organization* if it is “closed” and semi-self-maintaining.
- A set A is called *self-maintaining* if there exists a flux vector with the following properties: (a) $J_\rho > 0$ for all $\rho \in \mathfrak{R}_A$, (b) $J_\rho = 0$ for all $\rho \notin \mathfrak{R}_A$, (c) $(\mathbf{S}\mathbf{f})_x \geq 0$ for all $x \in A$.
- A set A is called an *organization* if it is “closed” and self-maintaining (Figure 3, right diagram).

The structural part (Section 2) of the theory of chemical organizations [12] is mostly concerned with lattice theoretic considerations, such as conditions under which the set of all organizations of a chemical universe (X, \mathfrak{R}) forms a lattice.

In order to demonstrate the relevance of chemical organizations, we generated two series of networks of increasing size, using the ToyChem model [5]. There, a chemical reaction network is generated by starting with a list of seed molecules and predicting the outcomes of the reactions between them, and repeating this process iteratively on the resulting molecules. For the first series, the complexity of the chemistry was increased at each step by adding new chemical species to the starting molecules, as opposed to the second series, where the molecular size threshold above which reactions were neglected was increased. Thus in the former case new chemical species are introduced without allowing new reactions between old chemical species at each step, whereas in the latter case new reactions are allowed at each step. We expect that an organization will be preserved when only new chemical species are introduced. Indeed, this is nicely illustrated by the first resulting series of sets of organizations, as seen in Figure 4 (right plot). In contrast to Figure 4 (left plot), the lattice of

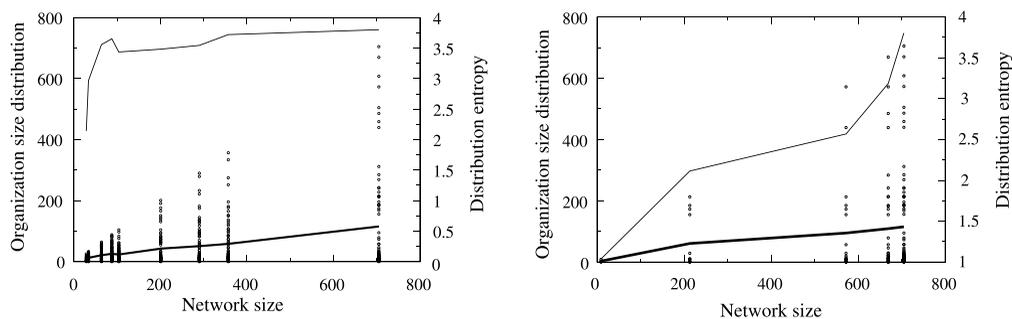


Figure 4. Change of the set of organizations in a growing network. Network size is increased by increasing the number of allowed molecules (left) or the number of starting molecules (right). For each network size, the sizes of all organizations form a distribution (\circ), which is shown with its mean (connected by the bold black line) and its entropy (connected by a light black line).

organizations of a smaller subnetwork is a sublattice of the lattice of organizations of the bigger network. Organization theory thus lets us study the modularity of a reaction network. Whereas some networks lead to the complex lattice shown in Figure 4, the purine synthesis network of [3] leads only to a set of organizations composed of the empty set and the complete network. This seems to indicate that the purine synthesis is not modular.

5 Set-Valued Set Functions

In [55] we considered the set of molecular types that can be produced by chemical reactions from subsets $A \in \mathcal{P}(X)$ as a function of the set A , with $\mathcal{P}(X)$ being the power set, that is, the set of all subsets of X . Using the notation introduced in the previous section, the function p' can be written as

$$p'(A) = \bigcup_{\rho \in \mathfrak{R}_A} \text{img} \rho. \quad (6)$$

A slightly more restrictive definition takes into account that a molecular species x may be effectively destroyed in a reaction ρ but still appears on the r.h.s. of that chemical equation. This is the case if $0 < \alpha_{x,\rho} < \bar{\alpha}_{x,\rho}$. Thus we set

$$\tilde{p}'(A) = \bigcup_{\rho \in \mathfrak{R}_A} \{x \in X \mid s_{x,\rho} > 0\}. \quad (7)$$

These functions, however, have the sometimes undesirable property of excluding inert species that only act as catalysts or that are not touched by any reaction, that is, those in the set

$$k(A) = \{x \in A \mid s_{x,\rho} = 0 \quad \forall \rho \in \mathfrak{R}_A\}. \quad (8)$$

In many situations it will be more natural to consider

$$p(A) = p'(A) \cup k(A) \quad \text{and} \quad \tilde{p}(A) = \tilde{p}'(A) \cup k(A). \quad (9)$$

Formally, p , p' , \tilde{p} , and \tilde{p}' are functions $\mathcal{P}(X) \rightarrow \mathcal{P}(X)$, so-called *set-valued set functions*. In a sense, these functions describe the structure of the chemical network on an aggregate level by the compound outcome of reaction mixtures (see Figure 5 for an example).

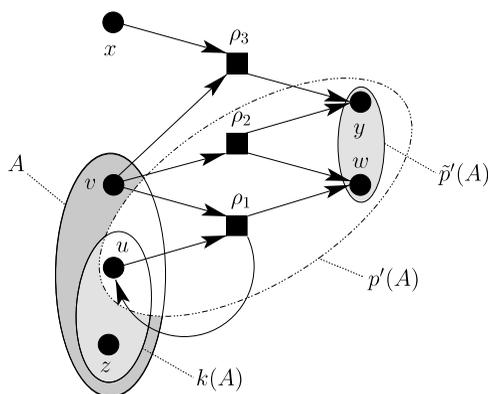


Figure 5. Illustration of the set-valued set functions k , p' , \tilde{p}' for the subset $A = \{u, v, z\}$ with u being a catalyst and z an inert species.

Comparing these definitions and the definition of “closed” and semi-self-maintaining sets, we find:

LEMMA 1: *A set A is “closed” if and only if $p'(\mathcal{A}) \subseteq A$. A set A is semi-self-maintaining if and only if $A \subseteq \tilde{p}(\mathcal{A})$.*

Proof: Recall that A is “closed” if and only if $\text{img}\rho \subseteq A$ whenever $\rho \in \mathfrak{R}_A$. Now suppose A is semi-self-maintaining. If $x \in A$ is consumed by some reaction, then it must be produced in some reaction, that is, $x \in \tilde{p}'(\mathcal{A})$. If x is not consumed by any reaction, then $s_{x,\rho} \geq 0$ for all $\rho \in \mathfrak{R}_A$, that is, either $x \in \mathcal{k}(\mathcal{A})$ or there is a reaction τ with $s_{x,\tau} > 0$, in which case $x \in \tilde{p}'(\mathcal{A})$. It follows that in each case $x \in \tilde{p}(\mathcal{A})$. Conversely, suppose $x \in \tilde{p}(\mathcal{A})$. Then $x \in \tilde{p}'(\mathcal{A})$ (i.e., x is produced by some reaction) or $x \in \mathcal{k}(\mathcal{A})$ (i.e., x is not consumed by any reaction). It follows that A is semi-self-maintaining.

Now, a semi-organization can be described as a fixed point of the set-valued set function \tilde{p} :

THEOREM 1: *A subset $A \subseteq X$ is a semi-organization of (X, \mathfrak{R}) if and only if $\tilde{p}(\mathcal{A}) = A$.*

Proof: Consider a semi-organization A . With Lemma 1, this is equivalent to $(p'(\mathcal{A}) \subseteq A \wedge A \subseteq \tilde{p}(\mathcal{A}))$, which is equivalent (cf. Equations 6 and 7) to $(\tilde{p}'(\mathcal{A}) \subseteq A \wedge A \subseteq \tilde{p}(\mathcal{A}))$, which, because $\mathcal{k}(\mathcal{A}) \subseteq A$, is equivalent to $(\tilde{p}'(\mathcal{A}) \cup \mathcal{k}(\mathcal{A}) \subseteq A \wedge A \subseteq \tilde{p}(\mathcal{A}))$, which is equivalent to $\tilde{p}(\mathcal{A}) = A$.

Lemma 1 suggests consideration of other set-valued set functions that might be used to characterize self-maintenance. We shall see that this approach leads in a rather natural way to a family of such functions.

We say $x \in A$ is *maintainable in A* if for each $\rho \in \mathfrak{R}_A$ with $x \in \text{dom}\rho$ there is a flux vector J such that (a) $J_\rho > 0$ and (b) $\hat{\mathbf{S}}_A J = \mathbf{0}$. In other words, x is maintainable in A if no reaction $\rho \in \mathfrak{R}_A$ inevitable leads to the depletion of x . In particular, x is maintainable in A if it is not required for any reaction of A at all. Let $m(\mathcal{A})$ denote the set of maintainable species in A .

LEMMA 2: *A set A is self-maintaining if and only if each $x \in A$ is maintainable in A , that is, if $m(\mathcal{A}) = A$.*

Proof: By taking a suitable positive (convex) combination we see that there is a flux vector J^x satisfying $\hat{\mathbf{S}}_A J^x = \mathbf{0}$ and $J_\rho^x > 0$ for all $\rho \in \mathfrak{R}_A$ with $x \in \text{dom}\rho$. Taking again a suitable positive convex combination, we can construct a stationary flux vector J^* (w.r.t. $\hat{\mathbf{S}}_A$) such that $J_\rho^* > 0$ for all $\rho \in \mathfrak{R}_{m(\mathcal{A})}$. Thus, $m(\mathcal{A}) = A$ is equivalent to the existence of a flux vector J for which (a) $J_\rho > 0$ for all $\rho \in \mathfrak{R}_A$ and (b) $(\mathbf{S}J)_x \geq 0$ for all $x \in A$, that is, to the fact that A is self-maintaining.

LEMMA 3: *For all $A \subseteq X$, $m(\mathcal{A}) \subseteq \tilde{p}(\mathcal{A})$ holds.*

Proof: If $x \in m(\mathcal{A})$, then there is a stationary flux vector satisfying $J_\rho > 0$ for all $\rho \in \mathfrak{R}_x := \{\rho \in \mathfrak{R}_A \mid x \in \text{dom}\rho\}$. If $\mathfrak{R}_x = \emptyset$ then $x \in \mathcal{k}(\mathcal{A})$. If this is not the case, then either $s_{x,\rho} = 0$ for all $\rho \in \mathfrak{R}_x$ (in which case $x \in \mathcal{k}(\mathcal{A})$ again), or there is at least one reaction ρ with $s_{x,\rho} \neq 0$. Since J is stationary with $J_\rho > 0$ for all ρ , in particular those with $s_{x,\rho} < 0$, it follows that there is at least one reaction $\tau \in \mathfrak{R}_A$ with $s_{x,\tau} > 0$ (otherwise J could not be stationary). In this last case $x \in \tilde{p}'(\mathcal{A})$.

The construction of $m(\mathcal{A})$ is somewhat unsatisfactory in that it does not distinguish between those species that can be produced from the set A and those that cannot be regenerated, for example, after a perturbation decreasing their concentration. Such a modified concept would be of particular interest, for example, in the context of metabolic networks, and is therefore made more precise in the following.

Let us say that $x \in X$ is *stationarily produced* from A if there is a stationary flux vector J (w.r.t. $\hat{\mathbf{S}}_A$) such that $J_{\psi_x} > 0$, that is, if the stationary flux produces an excess of x that we have to remove using the efflux reaction ψ_x in order to maintain balance. Note that we do not require each reaction with

$x \in \text{dom} \rho$ to have nonzero flux here. In fact, for example, in the context of metabolic networks it makes sense to assume that we can “switch off” certain reaction channels.

Now define $s(\mathcal{A})$ as the set of all species that are stationarily produced from \mathcal{A} . It is easy to show that there is a stationary flux vector that simultaneously produces all $x \in s(\mathcal{A})$: For each $x \in s(\mathcal{A})$, find a flux vector J_x that produces x , and take a strictly positive convex combination J^* of these vectors. Clearly J^* is again a stationary flux vector on \mathcal{A} , and by construction $J_{\psi_x} > 0$ for all $x \in s(\mathcal{A})$. The set of all such stationary flux vectors forms a cone \mathcal{K}_s .

By construction, $x \in s(\mathcal{A})$ implies that there is at least one reaction $\rho \in \mathfrak{R}_{\mathcal{A}}$ that produces x , that is, $s_{\rho,x} > 0$ for some $\rho \in \mathfrak{R}_{\mathcal{A}}$. Consequently, $s(\mathcal{A}) \subseteq \tilde{p}'(\mathcal{A})$.

An even stronger requirement on the flux vectors is to assume that there is inevitably a loss in all species that are produced. Formally, this means

$$J_{\psi_x} > 0 \quad \text{for all } x \in \bigcup_{\rho: J_{\rho} > 0} \text{img} \rho \tag{10}$$

Let $s^*(\mathcal{A})$ be the set of all $x \in X$ such that there is a stationary flux vector J on \mathcal{A} satisfying the conditions (10). By the same argument as above, there is a stationary flux vector J^* satisfying Equation 10 and $J_{\psi_x} > 0$ for all $x \in s^*(\mathcal{A})$, and an associated cone \mathcal{K}_{s^*} . By construction, $\mathcal{K}_{s^*} \subseteq \mathcal{K}_s$. This implies $s^*(\mathcal{A}) \subseteq s(\mathcal{A})$.

It is sometimes convenient to consider only those species x that are stationarily produced *within* a set \mathcal{A} , that is, $x \in \mathcal{A} \cap s(\mathcal{A})$. Clearly, any such species can also be maintained within \mathcal{A} . Thus we have $\mathcal{A} \cap s(\mathcal{A}) \subseteq m(\mathcal{A})$.

Let us summarize the mutual relationships of the set-functions defined so far (see Figure 6 for an example):

THEOREM 2: *For all $\mathcal{A} \subseteq X$ we have the following system of inclusions:*

$$\begin{array}{ccccccc} \mathcal{A} \cap s(\mathcal{A}) & \subseteq & m(\mathcal{A}) & \subseteq & \tilde{p}(\mathcal{A}) & \subseteq & p(\mathcal{A}) \\ & & & & \cup & & \cup \\ s^*(\mathcal{A}) & \subseteq & s(\mathcal{A}) & \subseteq & \tilde{p}'(\mathcal{A}) & \subseteq & p'(\mathcal{A}) \end{array}$$

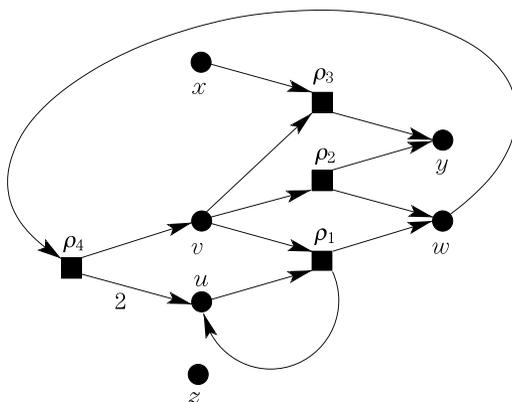


Figure 6. Illustration of the set-valued set functions m , s , and s^* for the set $\mathcal{A} = \{u, v, w, x, y, z\}$. The reaction network is essentially the one from Figure 5 with the additional reaction ρ_4 . The species x and v are not maintainable. The set of self-maintaining chemical species is $m(\mathcal{A}) = \{u, w, y, z\}$. If outflow is applied to the produced species in the set \mathcal{A} , the set of maintainable species is reduced to the set of stationarily produced species $s(\mathcal{A}) = \{u\}$. If outflow is applied to all species in \mathcal{A} , the set of stationarily produced species is reduced to s^* , which in this case is the empty set.

6 Isotonic and Non-isotonic Functions

One of the most basic (and desirable) properties of set-valued set functions is isotony: A function $u : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ is *isotonic* if $A \subseteq B$ implies $u(A) \subseteq u(B)$.

THEOREM 3: p' , \tilde{p}' , s , and s^* are isotonic functions.

Proof: It is easy to see that p' is isotonic. If $x \in p'(A)$, then there is a reaction ρ with $\text{dom}\rho \subseteq A$ and $x \in \text{img}\rho$. It follows that $\text{dom}\rho \subseteq B$ and hence $x \in p(B)$. An analogous argument works for \tilde{p}' .

In order to show the analogous results for s and s^* , we need to make sure that enlarging the subset A does not interfere with the stationary flux vectors. The following diagram shows the structure of the modified stoichiometric matrix $\hat{\mathbf{S}}_B$. The diagonal block at the right-hand side denotes the efflux reactions ψ_x for each $x \in X$ with entry -1 :

$$\hat{\mathbf{S}}_B = \begin{array}{c} \begin{array}{|c|} \hline \mathfrak{R}_B \\ \hline \end{array} \\ \begin{array}{|c|} \hline \mathfrak{R}_A \\ \hline \end{array} \\ \begin{array}{|c|} \hline X \\ \hline \end{array} \end{array} \begin{array}{|c|} \hline 0 \\ \hline \end{array} \quad (11)$$

We see that $\hat{\mathbf{S}}_A$ is a submatrix of $\hat{\mathbf{S}}_B$ with a very simple structure: The columns representing the non-efflux reactions in $\mathfrak{R}_B \setminus \mathfrak{R}_A$ are missing; otherwise the two matrices coincide. A solution of $\hat{\mathbf{S}}_A J = 0$ thus has a corresponding solution $\hat{\mathbf{S}}_B J' = 0$, where J' is obtained from J by adding 0 entries to the $\mathfrak{R}_B \setminus \mathfrak{R}_A$ coordinates; otherwise J and J' are identical. Clearly, we have $J_\rho > 0$ only for reactions with $\text{dom}\rho \in A \subset B$ (except for efflux reactions); thus J' is a stationary flux vector on B . Since J and J' coincide with respect to the efflux reactions, we have to conclude that $x \in s(A)$ implies $x \in S(B)$.

Furthermore, if J satisfies Equation 10, so does J' , and hence $x \in s^*(A)$ implies $x \in s^*(B)$.

In contrast, however, p , \tilde{p} , and m are *not* isotonic in general. As a counterexample, consider the following situation: Suppose there is $x \in A$ such that $x \notin \text{dom}\rho$ for all $\rho \in \mathfrak{R}_A$. For any flux vector J satisfying the conditions in the definition of m above, we have $(\mathbf{S}J)_x = 0$ and hence $x \in m(A) \subseteq \tilde{p}(A) \subseteq p(A)$. Assume, furthermore, that there is a single reaction $\theta \in \mathfrak{R}_B$ for some B satisfying $A \subset B$ that consumes x , that is, $x \in \text{dom}\theta$. Furthermore, suppose $x \notin \text{img}\theta$. Clearly, $x \notin p(B)$ and hence also $x \notin \tilde{p}(B)$, which in turn implies $x \notin m(B)$. It follows in particular that $p(A) \not\subseteq p(B)$, $\tilde{p}(A) \not\subseteq \tilde{p}(B)$, and $m(A) \not\subseteq m(B)$.

A reaction network (X, \mathfrak{R}) is a *flow system* if for each $x \in X$ there is an efflux reaction ψ_x . It follows immediately that $k(A) = \emptyset$ and $\hat{\mathbf{S}}_A = \mathbf{S}_A$ for all $A \in \mathcal{P}(X)$.

LEMMA 4: Let (X, \mathfrak{R}) be a flow system. Then $m : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ is isotonic.

Proof: By construction there is a flux vector J satisfying $\mathbf{S}_A J = 0$, $J_\rho > 0$ for all $\rho \in \mathfrak{R}_A$, and in particular $J_{\psi_x} > 0$ for all $x \in m(A)$. Now consider $m(B)$. Analogously, there is a flux vector J' satisfying $\mathbf{S}_B J' = 0$. Now construct a flux vector J^* such that $J^*_\rho = J_\rho + \epsilon J'_\rho$ for all $\rho \in \mathfrak{R}_B$ with the exception of the efflux reactions ψ_x with $x \in m(A)$. The flux through the latter is adjusted so that stationarity is maintained. For small enough ϵ , this can be achieved in such a way that $J^*_{\psi_x} > 0$. It follows that $m(A) \subseteq m(B)$.

THEOREM 4: *Let (X, \mathfrak{R}) be a flow system. Then \mathcal{A} is self-maintaining if and only if $\mathcal{A} \subseteq s(\mathcal{A})$.*

Proof: In a flow system, any stationary flux vector on \mathcal{A} satisfying $(\mathbf{S}_{\mathcal{A}}J)_x = 0$ for all $x \in \mathcal{A}$ and $J_{\psi_x}^* > 0$ can be modified also to satisfy $J_{\rho} > 0$ for all $\rho \in \mathfrak{R}_{\mathcal{A}}$ by reducing the effluxes J_{ψ_x} in return. Choosing J_{ρ} small enough, this can clearly be done in such a way that $J_{\psi_x} > 0$. Thus, if $\mathcal{A} \subseteq s(\mathcal{A})$ we have also $m(\mathcal{A}) = \mathcal{A}$. The converse is obvious.

We remark that this result also holds for flow systems with permanent molecules [12].

LEMMA 5: *Let (X, \mathfrak{R}) be a flow system. Then for all \mathcal{A} , $s(\mathcal{A}) \cap \mathcal{A} = m(\mathcal{A})$.*

Proof: In general, a stationarily produced molecule is also maintainable in \mathcal{A} , that is, $\mathcal{A} \cap s(\mathcal{A}) \subseteq m(\mathcal{A})$. It remains to be shown that in a flow system a maintainable molecule in \mathcal{A} is stationarily produced in \mathcal{A} : In a flow system there is an efflux reaction ψ_x for every molecule $x \in \mathcal{A}$. If a molecule is maintainable, this efflux reaction can be compensated (i.e., there is a stationary flux vector J with $J_{\psi_x} > 0$). Therefore the molecule is also stationarily produced.

7 Generalized Topologies

The connection of our exposition so far with the framework of (point set) topology is the fact that some meaningful topological concepts can already be defined on a set X endowed with an arbitrary set-valued set function. This section summarizes, and in part slightly extends, results by Day [11], Hammer [31, 26], and Gniłka [28] on such structures.

Let X be an arbitrary set, and let $\mathbf{cl} : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ be an arbitrary set-valued function. We shall see that it is fruitful to interpret \mathbf{cl} as a *closure* function on X ; hence we call $\mathbf{cl}(\mathcal{A})$ the *closure* of the set \mathcal{A} . The dual of the closure function is the *interior function* $\mathbf{int} : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ defined by

$$\mathbf{int}(\mathcal{A}) = X \setminus \mathbf{cl}(X \setminus \mathcal{A}). \quad (12)$$

Given the interior function, we obviously recover the closure as $\mathbf{cl}(\mathcal{A}) = X \setminus (\mathbf{int}(X \setminus \mathcal{A}))$. Note that the distinction of closure and interior is completely arbitrary in the absence of additional conditions.

The most immediate connection between chemical organizations and topological concepts is provided by the “closure” function C denoted in [12] by G_{CL} .

On the other hand, we might as well use any of the set-valued set functions defined in the previous sections to capture the topological structure of a chemical reaction network. We argue that the most natural choice is the closure function

$$\mathbf{cl}(\mathcal{A}) = p(\mathcal{A}) \cup \mathcal{A} = \tilde{p}(\mathcal{A}) \cup \mathcal{A} = p'(\mathcal{A}) \cup \mathcal{A} = \tilde{p}'(\mathcal{A}) \cup \mathcal{A}. \quad (13)$$

Indeed, $\mathbf{cl}(\mathcal{A})$ describes the state of the chemical system after the products obtainable from \mathcal{A} have been produced but before the original material has been consumed completely (e.g., Figure 3, second column of the table). It is easy to check that \mathbf{cl} is isotonic and enlarging.

The function \mathbf{cl} also has an easy relation with the “closure” $C(\mathcal{A})$: by construction $\mathcal{A} \subseteq C(\mathcal{A})$. Since $C(\mathcal{A})$ is closed, it must also contain everything that can be generated from \mathcal{A} , that is, $p(\mathcal{A}) \subseteq C(\mathcal{A})$, and hence $\mathbf{cl}(\mathcal{A}) \subseteq C(\mathcal{A})$. But this implies that $C(\mathcal{A}) = C(\mathbf{cl}(\mathcal{A}))$. Repeating the argument, we see $\mathbf{cl}^n(\mathcal{A}) \subseteq C(\mathcal{A})$ for all n . In the case of infinite sets, n may be any ordinal number. It can be proven that there is in general a smallest ordinal number η such that $\mathbf{cl}^{\eta}(\mathcal{A}) = C(\mathcal{A})$. The closure function \mathbf{cl} is implicitly used in the “generate closure” operator in [12].

In topological theory, a set $\mathcal{A} \in \mathcal{P}(X)$ is *closed* if $\mathcal{A} = \mathbf{cl}(\mathcal{A})$ and *open* if $\mathcal{A} = \mathbf{int}(\mathcal{A})$. Clearly, \mathcal{A} is closed if and only if $\mathcal{A} = C(\mathcal{A})$. A set \mathcal{A} is open if its complement $X \setminus \mathcal{A}$ is closed. In contrast

to classical topology, open and closed sets do not define the structure of our chemical universe. Nevertheless, these sets define interesting subsets [12, 15].

Alternatively, we could use $i(\mathcal{A}) = \hat{p}(\mathcal{A}) \cap \mathcal{A}$ as an interior function. (Note that $i(\mathcal{A})$ and $\text{cl}(\mathcal{A})$ are in general not duals.) In the resulting topological structure, semi-self-maintaining sets correspond to open sets. Repeated application of i eventually (after a finite number of steps in finite sets) leads to an idempotent interior function $I(\mathcal{A}) = i^n(\mathcal{A})$. In consistent reaction systems, $I(\mathcal{A})$ is the largest semi-self-maintaining set contained in \mathcal{A} . In this structure, semi-self-maintaining sets are exactly the open sets in (X, i) .

Both the idempotent closure C and the idempotent interior I are used explicitly in [12] as a means to generate “closed” and semi-self-maintaining sets. In fact, cl and i are used implicitly to computationally construct C and I , respectively.

Interestingly, any extended topology (isotonic space) can be seen as a combination of an expanding and a contracting isotonic function, that is, as a closure and an interior function defining two different neighborhood spaces:

LEMMA 6: Let $u : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ and $v : \mathcal{P}(X) \rightarrow \mathcal{P}(X)$ be two isotonic functions such that $u(\mathcal{A}) \subseteq \mathcal{A} \subseteq v(\mathcal{A})$, and let $w(\mathcal{A}) = u(\mathcal{A}) \cup [v(\mathcal{A}) \setminus \mathcal{A}]$. Then $u(\mathcal{A}) = w(\mathcal{A}) \cap \mathcal{A}$ and $v(\mathcal{A}) = w(\mathcal{A}) \cup \mathcal{A}$.

Proof: Direct computation.

Let cl and int be a closure function and its dual interior function on X . Then the *neighborhood function* $\mathcal{N} : X \rightarrow \mathcal{P}(\mathcal{P}(X))$ assigns to each point $x \in X$ the collections

$$\mathcal{N}(x) = \{N \in \mathcal{P}(X) \mid x \in \text{int}(N)\} \quad (14)$$

of its neighborhoods. Closure and neighborhood are equivalent [11]. More precisely:

$$x \in \text{cl}(\mathcal{A}) \iff (X \setminus \mathcal{A}) \notin \mathcal{N}(x) \quad \text{and} \quad x \in \text{int}(\mathcal{A}) \iff \mathcal{A} \in \mathcal{N}(x). \quad (15)$$

See Figure 7 for an example, and the Web Supplement for further details.

8 Connectedness

Topological connectedness is closely related to separation. Two sets $\mathcal{A}, B \in \mathcal{P}(X)$ are *semi-separated* if there are neighborhoods $N' \in \mathcal{N}(\mathcal{A})$ and $N'' \in \mathcal{N}(B)$ such that $\mathcal{A} \cap N'' = N' \cap B = \emptyset$; they are *separated* if there are neighborhoods $N' \in \mathcal{N}(\mathcal{A})$ and $N'' \in \mathcal{N}(B)$ such that $N' \cap N'' = \emptyset$.

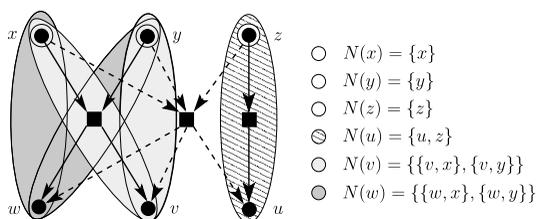


Figure 7. Minimal neighborhoods for each of the species $i \in \{x, y, z, u, v, w\}$ in a reaction network composed of two elementary reactions (bold arrows) and a super-reaction (dashed arrows) that performs the same chemical transformation as the superposition of the two elementary reactions. By definition, all supersets of neighborhoods are also neighborhoods. In contrast to graphs, the minimal neighborhoods are not unique. For instance, w has the two minimal neighborhoods $\{w, x\}$ and $\{w, y\}$.

A set $Z \in \mathcal{P}(X)$ is *connected* in a space (X, cl) if it is not a disjoint union of a nontrivial semi-separated pair of sets $A, Z \setminus A$ with $A \neq \emptyset, Z$.

For proofs of the statements in this paragraph we refer to the supplementary material of [55]. If (X, cl) is isotonic, then A and B are semi-separated if and only if $\text{cl}(A) \cap B = A \cap \text{cl}(B) = \emptyset$. Connectedness in isotonic spaces can thus be characterized by the Hausdorff-Lennes condition: A set $Z \in \mathcal{P}(X)$ is connected in an isotonic space (X, cl) if and only if for each proper subset $A \subseteq Z$ it holds that

$$[\text{cl}(A) \cap (Z \setminus A)] \cup [\text{cl}(Z \setminus A) \cap A] \neq \emptyset. \quad (16)$$

The collection of connected sets satisfies the following three properties in isotonic spaces [30]:

- (c1) If Z consists of a single point, then Z is connected.
- (c2) If Y and Z are connected and $Y \cap Z \neq \emptyset$, then $Y \cup Z$ is connected.
- (c3) If Z is connected, then $\text{cl}(Z)$ is also connected.

There have been several attempts to use connectedness as the primitive notion in topological theory [61, 32, 33].

This topological notion of connectedness nicely coincides with the notion of connectedness in directed graphs: In this case $\text{cl}(A)$ is the set of out-neighbors of the vertices $x \in A$ plus A itself. Thus A is connected unless there is a nontrivial bipartition A', A'' of A without an arrow pointing from A' to A'' or vice versa.

However, this notion has severe shortcomings in more general neighborhood spaces. Intuitively, one would like to consider subspaces of (X, ρ) that are spanned by a single reaction as connected sets. However, this is in general not the case. Consider $X = \{u, v, x, y\}$ with a single reaction $\rho : u + v \rightarrow x + y$. One easily can check that $A = \{u, x\}$ and $\bar{A} = \{v, y\}$ form a semi-separation, and hence X is not connected; see also Figure 8. Topological connectedness is thus too restrictive in the context of chemical reaction networks.

Another natural notion of connectedness derives from the hypergraph representation [56, 63] of (X, \mathfrak{R}) : We might say that a set is “graphically connected” if it is connected in the “substrate graph” of (X, \mathfrak{R}) [59], which is obtained by connecting two molecules x and y whenever they are involved in the same reaction ρ , that is, $\{x, y\}$ is an edge if and only if there is a reaction ρ such that $\{x, y\} \subseteq$

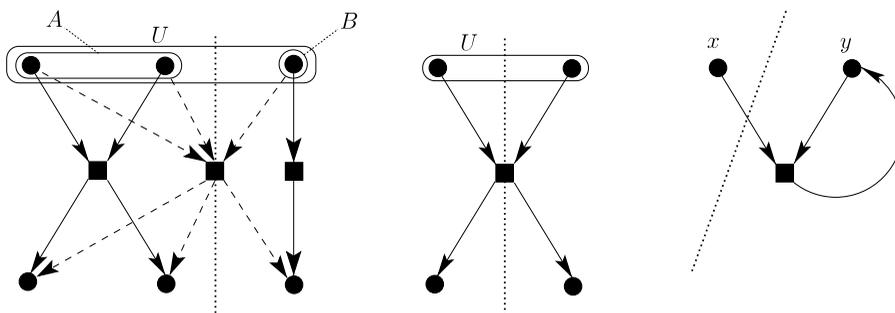


Figure 8. Illustration of the concepts separated, semi-separated, and productively connected. Left diagram: The reaction network is separated although from a graph theoretical viewpoint the network is connected. This classification makes sense in that in a chemical sense nothing new is gained by the super-reaction (dashed arrows) that holds the network together in the sense of graph theory. Middle diagram: The reaction network is semi-separated. Right diagram: The reaction network is productively connected with respect to p' , since something new is produced from the union of the subsets ($p'(\{x\}) = p'(\{y\}) = p'(\{x\}) \cup p'(\{y\}) = \emptyset$ and $p'(\{x\} \cup \{y\}) = \{y\}$). With respect to p the reaction network is not productively connected, because nothing new can be produced that was not already part of one of the two subsets ($p(\{x\}) = \{x\}$, $p(\{y\}) = \{y\}$, $p(\{x\}) \cup p(\{y\}) = \{x, y\}$, and $p(\{x\} \cup \{y\}) = \{y\}$).

$\text{dom}\rho \cup \text{img}\rho$. This notion of connectedness, however, does not really seem very useful, since it is easy to construct quite large graphically connected sets on which not a single reaction is possible. To see this, we consider the following construction. We start from

$$B_x = \bigcup_{\rho: x \in \text{dom}\rho} (\text{dom}\rho \cup \text{img}\rho) \quad (17)$$

and remove molecules $y \neq x$ until $\text{dom}\rho \not\subseteq B'$ for any ρ . By construction, each $y \in B'$ is connected to x in the substrate graph; thus B' is graphically connected. In general, B' can be a very large set.

As a middle ground we propose to consider a more restrictive notion of separation between sets. This concept then implies a less restrictive notion of connectedness.

We say that $A, B \in \mathcal{P}(X)$ are *productively separated* if for all $Z \subseteq A \cup B$ it holds that

1. $\text{cl}(Z \cap A) \cap B = \emptyset$ and $\text{cl}(Z \cap B) \cap A = \emptyset$,
2. $\text{cl}(Z) = \text{cl}(Z \cap A) \cup \text{cl}(Z \cap B)$.

If (X, cl) is an isotonic space, then A and B are semi-separated if condition 1 holds for all $Z \subseteq A \cup B$. (To see this, observe that isotony implies that condition 1 is true for all $Z \subseteq A \cup B$ whenever it is true for $Z = A \cup B$.) Condition 2 becomes trivial in additive spaces. In non-additive spaces, however, condition 2 encapsulates an important property: From two productively separated sets we cannot produce anything that we cannot obtain from the individual sets. In terms of our chemical networks, there are no reaction products that we can obtain only by mixing A and B .

LEMMA 7: *Suppose A and B are productively separated in an isotonic space (X, cl) . Let $A' \subseteq A$ and $B' \subseteq B$ be nonempty. Then A' and B' are also productively separated.*

Proof: Follows directly from the definition (choosing $Z = A' \cup B'$).

It is now natural to call a set Z *productively connected* if it cannot be decomposed into two nonempty subsets Z' and $Z'' = Z \setminus Z'$ with $Z' \cap Z'' = \emptyset$ that are productively separated. In general, if Z is connected, then it is also productively connected. In pretopological spaces (and in particular in digraphs), semi-separation and productive separation coincide; hence Z is productively connected if and only if it is connected in this case.

The following lemma shows that productive connectedness matches much closer our intuition of a “connected” reaction network. Indeed, most individual reactions correspond to connected sets:

LEMMA 8: *Consider a chemical universe $(X, \{\rho\})$ with $\text{dom}\rho \cup \text{img}\rho = X$ and $\text{dom}\rho \neq X$. Then $(X, \{\rho\})$ is productively connected w.r.t. the closure function p .*

Proof: Consider a proper split $\{A, \bar{A}\}$ of X , that is, $A \cup \bar{A} = X$, $A \cap \bar{A} = \emptyset$, and both A and \bar{A} are nonempty. We can distinguish two major cases: (i) $\text{dom}\rho \subseteq A$, and (ii) $\text{img}\rho \setminus A$ and $A \setminus \text{img}\rho$ are both nonempty. The third case, $A \cap \text{img}\rho = \emptyset$, reduces to case (i) by exchanging A and \bar{A} .

Case (i): We have $X = p(\text{dom}\rho) \subseteq p(A)$, that is, $p(A) \cap \bar{A} \neq \emptyset$, and hence the split $\{A, \bar{A}\}$ is not a semi-separation of X .

Case (ii): We have $p(A) = A$ and $p(\bar{A}) = \bar{A}$, that is, the split $\{A, \bar{A}\}$ is a semi-separation. There is, however, $Z = \text{dom}\rho$ such that $p(Z \cap A) = Z \cap A$, $p(Z \cap \bar{A}) = Z \cap \bar{A}$, and hence $p(Z \cap A) \cup p(Z \cap \bar{A}) = Z$. On the other hand, $p(Z) = X$. By assumption, $Z \neq X$. Thus the split $\{A, \bar{A}\}$ is not a productive semi-separation of X .

It follows that $(X, \{\rho\})$ does not admit a productive semi-separation w.r.t. p ; hence it is productively connected w.r.t. to p .

Note that $(X, \{\rho\})$ is in general not connected w.r.t p if $\text{img}\rho \subseteq \text{dom}\rho = X$ (Figure 8).

It seems natural to call a reaction *topologically elementary* if (i) $\text{img}\rho \not\subseteq \text{img}\rho$ and (ii) there is no other reaction ρ' with $\text{dom}\rho' \cup \text{img}\rho' \subseteq \text{dom}\rho \cup \text{img}\rho$ with the exception of the reverse reaction $\rho' = \bar{\rho}$. The following theorem shows that productively connected sets can be built up from overlapping topologically elementary reactions. It will be an interesting question to characterize the connected sets that are spanned by individual topologically non-elementary reactions.

THEOREM 5: *In neighborhood spaces, productive connectedness has the following properties:*

- (c1) *If Z consists of a single point, then Z is productively connected.*
- (c2) *If Y and Z are productively connected and $Y \cap Z \neq \emptyset$, then $Y \cup Z$ is connected.*
- (c3) *If Z is productively connected, then $\text{cl}(Z)$ is also productively connected.*
- (c4) *Let $\{Z_i\}_{i \in I}$ be an arbitrary collection of connected sets such that $\bigcap_{i \in I} Z_i \neq \emptyset$. Then $W := \bigcup_{i \in I} Z_i$ is connected.*

Proof: Property (c1) is trivial, since there is no nontrivial partition.

Suppose that $Y \cup Z$ can be decomposed into two nonempty productively separated sets P and Q . By Lemma 7, $P \cap Y$ and $Q \cap Y$ are also productively separated unless one of the intersections is empty. Since Y is assumed to be productively connected, we have to assume that one of the intersections is empty. The same argument can be made for Z . Thus the only possibility is $Y \subseteq P$ and $Z \subseteq Q$ (or vice versa). On the other hand, we have $Y \cap Z = \emptyset$. It follows that $Y \cup Z$ must be productively connected.

A similar argument shows that $\text{cl}(Y)$ is connected. Again, suppose $\text{cl}(A) = P \cup Q$ for two productively separated sets P and Q . In a neighborhood space, $Y \subseteq \text{cl}(Y)$. Thus $P \cap Y$ and $Q \cap Y$ are semi-separated unless one of them is empty. It follows that either $Y \subseteq P$ or $Y \subseteq Q$. But then either $\text{cl}(P)$ or $\text{cl}(Q)$ contains $\text{cl}(Y)$ by isotony, that is, either Q or P is empty, a contradiction. Hence $\text{cl}(Y)$ is productively connected.

Suppose W is not connected. Then there is a productive separation $W = W' \dot{\cup} W''$ with nonempty sets W' and W'' . Observe that in this case $Z'_i = Z_i \cap W'$ and $Z''_i = Z_i \cap W''$ for a productive separation of Z_i , unless either Z'_i or Z''_i is empty. By construction, there is a point $z_i \in Z_i$ for all $i \in I$. W.l.o.g. we may assume $z_i \in W'$. This implies $Z''_i = \emptyset$ and hence $Z_i \subseteq W'$ for all i , and hence $W'' = \emptyset$, contradicting the assumption that W admits a productive separation. Therefore, W is connected.

It follows immediately that for each set A and each point $x \in A$ the connected component $A[x]$ of A that contains x is well defined:

$$A[x] = \bigcup \{A' \subset A \mid x \in A' \text{ and } A' \text{ is connected}\}. \quad (18)$$

We see that productive connectedness provides a generalization of the usual notion of connectedness in neighborhood spaces that has the same convenient mathematical properties and a more intuitive interpretation. Since the two concepts coincide for the much better studied pretopological and topological spaces, one might argue that productive connectedness is the most natural notion of connectedness in neighborhood spaces.

9 Conclusion and Outlook

We have shown here how the stoichiometric picture of a chemical reaction network can be interpreted within the framework of point set topology. Specifically, we consider a set of molecular

species to be a (generalized) topological space, whose structure is determined by mutual accessibility via chemical reactions. The translation of concepts from chemical reaction network theory to this much more abstract mathematical framework gives access to a rich “language” and a collection of concepts—including connectedness (a focus in this study), compactness, convergence, regularity, and uniformity—that have been developed and explored by mathematicians for more than a century. We have demonstrated here in some detail how these topological notions can be translated back to the concrete case of chemical reaction networks. As our understanding of structural properties of large-scale chemical networks is still in its infancy, it remains an interesting research agenda to explore the chemical meaning of these abstract concepts in chemical reaction network theory. The same basic mathematical theory describes, for example, search spaces in molecular evolution as well as in combinatorial optimization settings that use complex search operators such as crossover [54].

The basic building blocks of the topological approach are set-valued set functions that describe which molecules can be generated from a given seed set of compounds. These functions implicitly arrange the chemical universe in a way that places molecules close to each other if they can be interconverted by applying few elementary reactions. Chemical organization theory [12] naturally can be reformulated in the topological language, with organizations taking on the role of appropriately defined closed sets.

The reformulation of chemical organization theory in topological terms and the theorems derived here have also practical implications. The definition of self-maintenance in terms of a set-valued set function highlights algebraic properties of this function, which in turn suggest novel avenues for algorithmic improvements. For instance, our results (Section 5, Lemma 2) show that self-maintenance can be verified by checking each molecule separately to determine whether it is maintainable, instead of trying to find a flux vector that maintains all molecules at once. Even if this does not lead to faster algorithms, we might use the concepts of “maintainable” and “stationarily produced” as a measure of the degree of maintainability of a set of molecules. For instance, the relative size of the set of stationarily produced molecules can be taken as a structural measure of robustness of an organization, because stationarily produced molecules can be regenerated when destroyed. Furthermore, we can take the new concepts to explain why a set is not self-maintaining, by identifying those molecules that are not maintainable.

The topological notions of neighborhood, separateness, and productively connectedness are rather abstract and yet difficult to comprehend intuitively from a biochemist’s point of view. Nevertheless we have shown that they lead to rather natural or at least interesting characterizations of the structure of a network. A challenging task will be to find their right interpretations and application range, which will be a prerequisite for making the theory accessible through software tools.

Network comparison and network alignment are interesting potential applications of our approach. Forst et al. [24] have shown that the extension of set algebra to chemical reaction networks opens the route to the identification of distinct metabolic features in two sets of organisms. Yet it is unclear if this also has interesting consequences in the topological framework.

The topological interpretation of chemical reaction networks and their organizations emphasizes the distinction between structural and dynamical properties, since dynamics has to be understood as a process living on the underlying topological structures. Exploring the relation of structural features of the network to dynamical properties—as has been done for chemical organizations ([12], Theorem 1)—has been identified as an important element of future research. For instance, Feinberg and collaborators [9] recently characterized those mass-action networks that can support bistable behavior on the basis of structural features of the “species reaction graph.” It will be interesting to see if their results have a natural topological interpretation as well.

As sketched above, there are various open problems, which cumulate into the fundamental question of how topological properties are related to dynamical behavior or, vice versa, how observed dynamics can be explained in terms of local topological properties of the network.

For some of the formal concepts and mathematical theorems discussed in this contribution it is not yet clear to what extent they are useful in practice to analyze and engineer chemical reaction

networks. At the very least, however, our work is useful in its explaining of network concepts in a well-established mathematical framework and hence forming a solid basis for further work, including the development of new or improved algorithmic approaches to analyze the structure of large chemical networks.

Acknowledgments

This work owes its existence to a number of free-form environments far away from office desks, computers, telephones, and e-mails: the beaches of Brighton (England), Ulcin (Montenegro), Oahu, and Hawai'i, and the winter landscape at Lake Bled in Slovenia. We thank Peter Wills and the anonymous reviewers for their comments on the manuscript. Financial support by the VW Foundation Grant "Evolution of networks: robustness, complexity and adaptability" to P. F. S., the DFG Graduiertenkolleg Wissensrepräsentation to G. B., and the BMBF (grant 0312704A) to F. C. and P. D. is gratefully acknowledged.

References

1. Bagley, R. J., & Farmer, J. D. (1992). Spontaneous emergence of a metabolism. In C. G. Langton, C. Taylor, J. D. Farmer, & S. Rasmussen (Eds.), *Artificial life II*, Santa Fe Institute Studies in the Sciences of Complexity (pp. 93–141). Redwood City, CA: Addison-Wesley.
2. Banzhaf, W., Dittrich, P., & Eller, B. (1999). Self-organization in a system of binary strings with spatial interactions. *Physica D*, *125*, 85–104.
3. Benkő, G. (2006). *Uncovering the structure of an artificial chemistry universe*. Ph.D. thesis, Faculty for Mathematics and Computer Science, University of Leipzig.
4. Benkő, G., Flamm, C., & Stadler, P. F. (2003). Generic properties of chemical networks: Artificial chemistry based on graph rewriting. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, & J. Ziegler (Eds.), *Advances in artificial life* (pp. 10–20). Heidelberg: Springer-Verlag.
5. Benkő, G., Flamm, C., & Stadler, P. F. (2003). A graph-based toy model of chemistry. *Journal of Chemical Information and Computer Sciences*, *43*, 1085–1093.
6. Berry, G., & Boudol, G. (1992). The chemical abstract machine. *Theoretical Computer Science*, *96*, 217–248.
7. Centler, F., & Dittrich, P. (2007). Chemical organizations in atmospheric photochemistries: A new method to analyze chemical reaction networks. *Planetary and Space Science*, *55*(4), 413–428.
8. Clarke, B. L. (1988). Stoichiometric network analysis. *Cell Biophysics*, *12*, 237–253.
9. Craciun, G., Tang, Y., & Feinberg, M. (2006). Understanding bistability in complex enzyme-driven reaction networks. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 8697–8702.
10. Cupal, J., Kopp, S., & Stadler, P. F. (2000). RNA shape space topology. *Artificial Life*, *6*, 3–23.
11. Day, M. M. (1944). Convergence, closure, and neighborhoods. *Duke Mathematical Journal*, *11*, 181–199.
12. Dittrich, P., & Speroni di Fenizio, P. (2007). Chemical organization theory. *Bulletin of Mathematical Biology*, *69*(4), 1199–1231.
13. Dittrich, P., Ziegler, J., & Banzhaf, W. (2001). Artificial chemistries—A review. *Artificial Life*, *7*, 225–275.
14. Dugundji, J., & Ugi, I. (1973). Theory of the *be*- and *r*-matrices. *Topics in Current Chemistry*, *39*, 19–29.
15. Ebenhöf, O., Handorf, T., & Heinrich, R. (2004). Structural analysis of expanding metabolic network. *Genome Informatics*, *15*(1), 35–45.
16. Feinberg, M. (1972). Complex balancing in general kinetic systems. *Archive for Rational Mechanics and Analysis*, *49*, 187–194.
17. Feinberg, M., & Horn, F. J. M. (1973). Dynamics of open chemical systems and the algebraic structure of the underlying reaction network. *Chemical Engineering Science*, *29*(3), 775–787.
18. Fell, D. A. (1997). *Understanding the control of metabolism*. London: Portland Press.
19. Fontana, W., & Buss, L. W. (1994). "The arrival of the fittest": Towards a theory of biological organisation. *Bulletin of Mathematical Biology*, *56*(1), 1–64.

20. Fontana, W., & Buss, L. W. (1994). What would be conserved if “the tape were played twice”? *Proceedings of the National Academy of Sciences of the United States of America*, *91*, 757–761.
21. Fontana, W., & Schuster, P. (1998). Continuity in evolution: On the nature of transitions. *Science*, *280*, 1451–1455.
22. Fontana, W., & Schuster, P. (1998). Shaping space: The possible and the attainable in RNA genotype-phenotype mapping. *Journal of Theoretical Biology*, *194*, 491–515.
23. Fontana, W. (1992). Algorithmic chemistry. In C. G. Langton, C. Taylor, J. D. Farmer, & S. Rasmussen (Eds.), *Artificial life II* (pp. 159–210). Redwood City, CA: Addison-Wesley.
24. Forst, C. V., Flamm, C., Hofacker, I. L., & Stadler, P. F. (2006). Algebraic comparison of metabolic networks, phylogenetic inference, and metabolic innovation. *BMC Bioinformatics*, *7*, 67.
25. Gagneur, J., & Klamt, S. (2004). Computation of elementary modes: A unifying framework and the new binary approach. *BMC Bioinformatics*, *5*, 175 [epub].
26. Gastl, G. C., & Hammer, P. C. (1967). Extended topology. Neighborhoods and convergents. In W. Fenichel (Ed.), *Proceedings of the Colloquium on Convexity 1965* (pp. 104–116). Copenhagen: Københavns Univ. Matematiske Inst.
27. Gleiss, P. M., Stadler, P. F., Wagner, A., & Fell, D. A. (2001). Relevant cycles in chemical reaction network. *Advances in Complex Systems*, *4*, 207–226.
28. Gniłka, S. (1994). On extended topologies. I: Closure operators. *Annales Societatis Mathematicae Polonae, Series I, Commentationes Mathematicae*, *34*, 81–94.
29. Guberman, J. M. (2003). Mass action networks and the deficiency zero theorem. Master’s thesis, Department of Mathematics, Harvard University.
30. Habil, E. D., & Elzenati, K. A. (2006). Connectedness in isotonic spaces. *Turkish Journal of Mathematics*, *30*, 247–262.
31. Hammer, P. C. (1962). Extended topology: Set-valued set functions. *Nieuw Archief voor Wiskunde III*, *10*, 55–77.
32. Hammer, P. C. (1963). Extended topology: Connected sets and Wallace separations. *Portugaliae Mathematica*, *22*, 77–93.
33. Harris, J. M. (2005). *Continuity and separation for point-wise symmetric isotonic closure functions* (Technical report). arXiv:math.GN/0507230.
34. Heinrich, R., & Rapoport, T. A. (1974). A linear steady-state treatment of enzymatic chains. General properties, control and effector strengths. *European Journal of Biochemistry*, *42*, 97–105.
35. Heinrich, R., & Schuster, S. (1996). *The regulation of cellular systems*. Chapman and Hall.
36. Hofbauer, J., & Sigmund, K. (1988). *Dynamical systems and the theory of evolution*. Cambridge, UK: Cambridge University Press.
37. Höllering, R., Gasteiger, J., Steinhauer, L., Schulz, K., & Herwig, A. (2000). The simulation of organic reactions: From the degradation of chemicals to combinatorial synthesis. *Journal of Chemical Information and Computer Sciences*, *40*, 482–494.
38. Hordijk, W., & Fontanari, J. F. (2003). Catalytic reaction sets, decay, and the preservation of information. *Proceedings of the IEEE International Conference on Integration of Knowledge Intensive Multi-Agent Systems* (pp. 133–138).
39. Hordijk, W., & Steel, M. (2004). Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, *227*, 451–461.
40. Horn, F. (1972). Necessary and sufficient conditions for complex balancing in chemical kinetics. *Archive for Rational Mechanics and Analysis*, *49*, 172–186.
41. Horn, F., & Jackson, R. (1972). General mass action kinetics. *Archive for Rational Mechanics and Analysis*, *B47*, 81–116.
42. Kacser, H., & Burns, J. A. (1973). The control of flux. *Symposia of the Society for Experimental Biology*, *32*, 65–104.
43. Kacser, H., Burns, J. A., & Fell, D. A. (1995). The control of flux: 21 years on. *Biochemical Society Transactions*, *23*, 341–366.

44. Kauffman, S. A. (1993). *The origin of order*. New York, Oxford: Oxford University Press.
45. McCaskill, J. S., & Niemann, U. (2000). Graph replacement chemistry for DNA processing. In A. Condon & G. Rozenberg (Eds.), *DNA computing* (pp. 103–116). Berlin: Springer.
46. Olivera, J. S., Bailey, C. G., Jones-Olivera, J. B., & Dixon, D. A. (2001). An algebraic-combinatorial model for the identification and mapping of biochemical pathways. *Bulletin of Mathematical Biology*, *63*, 1163–1196.
47. Petri, C. A. (1962). *Kommunikation mit Automaten*. Ph.D. thesis, University of Bonn, Bonn.
48. Schilling, C. H., & Palsson, B. Ø. (1998). The underlying pathway structure of biochemical reaction networks. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 4193–4198.
49. Schilling, C. H., Letscher, D., & Palsson, B. Ø. (2000). Theory for the systematic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology*, *203*, 229–248.
50. Schuster, S., Fell, D. A., & Dandekar, T. (2000). A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, *18*, 326–332.
51. Sellers, P. H. (1971, 1972). An introduction to a mathematical theory of chemical reaction networks I, II. *Archive for Rational Mechanics and Analysis*, *44*, 23–40, 376–386.
52. Speroni di Fenizio, P. (2000). A less abstract artificial chemistry. In M. Bedau, J. McCaskill, N. Packard, & S. Rasmussen (Eds.), *Artificial life VII* (pp. 49–53). Cambridge, MA: MIT Press.
53. Stadler, B. M. R., Stadler, P. F., Shpak, M., & Wagner, G. P. (2002). Recombination spaces, metrics, and pretopologies. *Zeitschrift für Physikalische Chemie*, *216*, 217–234.
54. Stadler, B. M. R., Stadler, P. F., Wagner, G. P., & Fontana, W. (2001). The topology of the possible: Formal spaces underlying patterns of evolutionary change. *Journal of Theoretical Biology*, *213*, 241–274.
55. Stadler, B. M. R., & Stadler, P. F. (2002). Generalized topological spaces in evolutionary theory and combinatorial chemistry. *Journal of Chemical Information and Computer Sciences*, *42*, 577–585.
56. Temkin, O. N., Zeigarnik, A. V., & Bonchev, D. G. (1996). *Chemical reaction networks: A graph-theoretical approach*. Boca Raton, FL: CRC Press.
57. Thürk, M. (1993). *Ein Modell zur Selbstorganisation von Automatenalgorithmen zum Studium molekularer Evolution*. Ph.D. thesis, Universität Jena, Germany.
58. Ugi, I., Stein, N., Knauer, M., Gruber, B., Bley, K., & Weidinger, R. (1993). New elements in the representation of the logical structure of chemistry by qualitative mathematical models and corresponding data structures. *Topics in Current Chemistry*, *166*, 199–233.
59. Wagner, A., & Fell, D. A. (2001). The small world inside large metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, *268*, 1803–1810.
60. Wagner, G., & Stadler, P. F. (2003). Quasi-independence, homology and the unity of type: A topological theory of characters. *Journal of Theoretical Biology*, *220*, 505–527.
61. Wallace, A. D. (1941). Separation spaces. *Annals of Mathematics*, *43*, 687–697.
62. Yung, Y. L., & DeMore, W. B. (1999). *Photochemistry of planetary atmospheres*. New York: Oxford University Press.
63. Zeigarnik, A. V. (2000). On hypercycles and hypercircuits in hypergraphs. In P. Hansen, P. W. Fowler, & M. Zheng (Eds.), *Discrete mathematical chemistry* (pp. 377–383). Providence, RI: American Mathematical Society.