

# GENERALIZED TOPOLOGIES: HYPERGRAPHS, CHEMICAL REACTIONS, AND BIOLOGICAL EVOLUTION

Christoph Flamm<sup>1</sup>, Bärbel M. R. Stadler<sup>2</sup>, and Peter F. Stadler<sup>1–6</sup>

<sup>1</sup>*Inst. f. Theoretical Chemistry, University of Vienna, Währingerstraße 17,  
A-1090 Vienna, Austria;*

<sup>2</sup>*Max Planck Institute for Mathematics in the Sciences, Inselstraße 22,  
D-04103 Leipzig, Germany;*

<sup>3</sup>*Bioinformatics Group, Department of Computer Science, and  
Interdisciplinary Center of Bioinformatics, Univ. Leipzig, Härtelstraße  
16-18, D-04107 Leipzig, Germany;*

<sup>4</sup>*RNomics Group, Fraunhofer IZI, Perlickstraße 1, D-04103 Leipzig,  
Germany;*

<sup>5</sup>*Center for non-coding RNA in Technology and Health, University of  
Copenhagen, Grønnegårdsvej 3, DK-1870 Frederiksberg, Denmark;*

<sup>6</sup>*The Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, New Mexico, USA*

**ABSTRACT:** In the analysis of complex networks, the description of evolutionary processes, or investigations into dynamics on fitness or energy landscapes notions such as similarity, neighborhood, connectedness, or continuity of change appear in a natural way. These concepts are of an inherently topological nature. Nevertheless, the connection to the mathematical discipline of point set topology is rarely made in the literature, presumably because in most applications there is no natural object corresponding to an open or closed set. The link to textbook topology thus cannot be made in a straightforward manner. Many of the deep results of point set topology still remain valid, however, when open sets are abandoned and generalizations of the closure operator are used as the foundation of the mathematical theory. Here we survey some applications of such generalized point set topologies to chemistry and biology, providing an overview of the underlying mathematical structures.

## INTRODUCTION

Investigations into general principles underlying chemical reaction networks, molecular phylogenetics, evolutionary changes, or the folding of biopolymers naturally encounter notions such as *similarity*, *proximity*, *connectedness*, or *continuity of change*. To a mathematician, all of these concepts are intrinsically topological in nature. On the other hand, all these systems are very far away from the “continuum” picture so familiar in physics and chemistry that familiar constructions such as  $\varepsilon$ -balls seem to be little use. Even worse, *the* starting point of textbook topology, the *open set*, does not have a natural counterpart in the paradigmatic examples that we will consider throughout this contribution.

Eduard Čech’s treatise of Topological Spaces [9], however, showed that the classical theory of Point Set Topology can be constructed in a different way, starting from Kuratowski’s axioms of closure functions or an equivalent notion of neighborhoods. As we shall see below, these can be seen as formalizations of a vague notion of accessibility that lies at the heart of many of the questions that we are interested in. Before proceeding to the mathematical framework, we briefly introduce a few of our model systems. We note, finally, that generalized topological spaces have applications in various applied domains of computer science, including digital image processing, information representation, the semantics of modal logic, or handwriting recognition [21, 66, 62, 31, 59, 56, 22, 60].

**Chemical Reaction Networks.** We start from a fixed set  $X$  of molecular types or chemical species. A chemical reaction  $\rho$  is a transformation rule of the form



where  $E_{\rho}^- := \{x \in X | a_{\rho,x}^- > 0\}$  are the educts and  $E_{\rho}^+ := \{x \in X | a_{\rho,x}^+ > 0\}$  are the products of reaction  $\rho$ . Thus we can interpret reaction  $\rho$  as a hyperedge  $\mathcal{E}_{\rho} = (E_{\rho}^-, E_{\rho}^+)$  in a directed hypergraph  $\mathfrak{H}$ . It may be convenient, even, to interpret  $E^+$  and  $E^-$  as multi-sets to incorporate the stoichiometric coefficients  $s_{\rho,x}^{\pm}$  as multiplicities [82]. The coefficients  $s_{\rho,x}^{\pm}$  are connected to the *stoichiometric matrix*  $\mathbf{S}$  via  $S_{\rho,x} = s_{+\rho,x} - s_{-\rho,x}$ . Figure 1 shows a graphical description of small chemical reaction network.

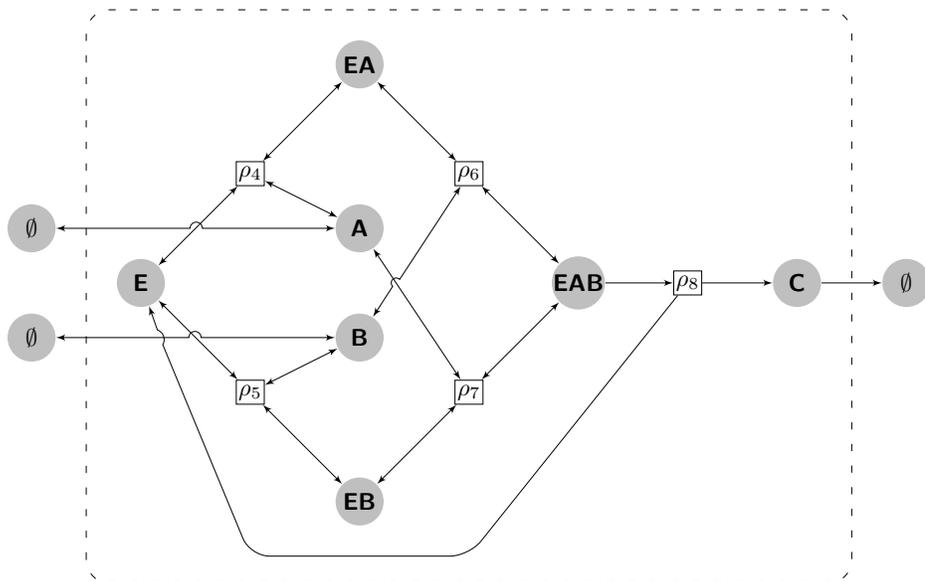


Figure 1: Hypergraph representation of the reaction network of elementary reaction steps for the overall enzyme catalyzed reaction  $A + B \xrightarrow{E} C$  with unordered substrate binding. Reaction nodes are drawn as squares and species nodes as gray circles. (Note for clarity the in- and outflow reaction nodes ( $\rho_1 - \rho_3$ ) have been omitted.)

Recently, the Network of Organic Chemistry (NOC) [23, 6] has been constructed from all organic reactions reported in the chemical literature since 1779. Despite the apparent complexity of this network, which captures all the knowledge on synthetic organic chemistry, NOC shows an unexpected but well-defined topological structure [37]. In particular, NOC possesses a densely wired core region of about 300 synthetically important building blocks and industrial compounds which are mutually connected by short synthetic pathways. The core region is embedded in a large and sparse periphery constituting approximately 78% of chemistry. Compounds in the periphery can be reached on average in 3-7 synthetic steps from the core region. Core region and periphery are surrounded by small isolated islands constituting about 18% of NOC. Much smaller chemical reaction networks that represent the metabolic pathways of several organism are compiled in dedicated databases, see e.g. [52].

In the analysis of such chemical networks, so-called *pathways*, i.e., sequences of reactions/hyperedges that connect chemical input with output molecules, and catalytic cycles are of particular interest. Chemical organizations, that is, closed and self-maintaining subsets of  $\mathfrak{S}$  [27, 18], furthermore, are closely related to the limit set of the corresponding reaction kinetics [61]. The most useful of these structural features are related to algebraic invariants that can be expressed in terms of  $\mathbf{S}$ , see [4] for a recent review.

The topological description of chemical networks centers describing what can be produced instantaneously from set  $A \subseteq X$  of molecular species, i.e.,

$$p(A) = \bigcup_{\rho: E_{\rho}^{-} \subseteq A} \{x \in X | s_{x,\rho} > 0\} \quad (2)$$

A species  $x$  is *maintainable* in  $A$  if no reaction  $\rho$  with  $E_{\rho}^{-} \subseteq A$  inevitably leads to the depletion of  $x$ . The set of maintainable species in  $A$  will be denoted by  $m(A)$ . A set  $A$  is closed if  $p(A) = A$  and self-maintaining if  $m(A) = A$ . So-called *flow systems* make the additional assumption that every species  $x$  can flow out of the system. Both  $p$  and  $m$  are set-valued set functions, i.e., they are of the form  $p: \mathfrak{P}(X) \rightarrow \mathfrak{P}(X)$ , where  $\mathfrak{P}(X)$  denotes the power-set of  $X$ . One can show that in flow systems both  $p$  and  $m$  are isotonic (see below), and hence impose a generalized topological structure on  $X$  [5].

**Evolutionary Computation and Genetic Evolution.** A (combinatorial) optimization problem is usually specified in terms of a set  $X$  of configurations and a cost function  $f: X \rightarrow \mathbf{R}$ , where  $\mathbf{R}$  is an ordered set, or, in the case of multi-objective optimization [17], a partially ordered set [74]. A large class of heuristic algorithms, including Simulated Annealing, Genetic Algorithms, Evolutionary Strategies, or Genetic Programming, attempt to find optimal solutions by moving through the set  $X$  and evaluating the cost function at different points  $x \in X$ . This search procedure imposes an implicit mathematical structure on the set  $X$  that determines how points or, more generally, subsets are mutually accessible. In a more biologically inspired setting, this *search space* is uniquely determined by the genetic operators, i.e., by mutation, recombination, duplication, deletion, or rearrangement of gene order.

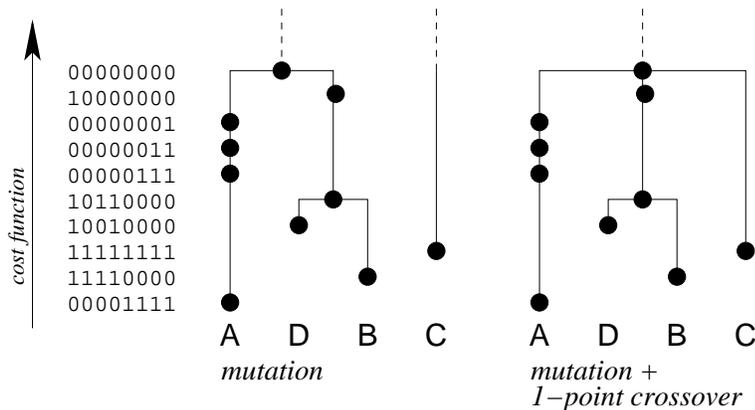


Figure 2: Barrier tree of a simple landscape with mutation (left) and mutation plus 1-point-crossover. Figure adapted from [26].

A natural way of abstracting the action of these operators is to determine for each “population”  $A \subseteq X$ , the set  $c(A)$  of configurations that can be reached from  $A$  by applying a single on of these operator. This is most easily visualized for mutations: Each parent  $x$  may give rise to a set  $c(\{x\})$  of possible offsprings (mutants). In this case,  $(X, c)$  defines a (possibly directed) graph. The situation becomes more complicated, however, when recombination (crossover) is considered [33]. The analogue of the adjacency relation of the graph is the recombination set  $\mathcal{R}(x, y)$ , which is defined as the set of all (possible) recombinants of two parents  $x$  and  $y$ . Recombination sets are usually required to satisfy (1)  $\{x, y\} \in \mathcal{R}(x, y)$ , and (2)  $\mathcal{R}(x, y) = \mathcal{R}(y, x)$ , Often (3)  $\mathcal{R}(x, x) = \{x\}$  is assumed, which is, however, not satisfied by models of unequal crossover [65, 72]. Functions  $\mathcal{R} : X \times X \rightarrow \mathcal{P}(X)$  satisfying these three axioms were also considered under different names, e.g. *transit functions* [10] and as *P-structures* [77, 75]. We note, furthermore, that recombination can be seen as a ternary relation  $(x, v, y) \in \widehat{\mathcal{R}} \iff v \in \mathcal{R}(x, y)$  closely connected with betweenness relations [1].

Similar to the chemical networks, we may base a formal treatment of topological structures on a set-values set-function that encodes reachability

at the level of sets by setting

$$c(A) = \bigcup_{x,y \in A} \mathcal{R}(x,y) \quad (3)$$

The topological structure defined by  $c$  also brings with it a concept of connectedness [81, 43], and hence allows a construction of level sets or “basins” for the cost function  $f : X \rightarrow \mathbf{R}$  as connected components (w.r.t.  $c$ ) of the sets  $X_h = \{x | f(x) \leq h\}$ . The basins have a hierarchical structure: if  $A$  and  $B$  are connected components of  $X_{h'}$  and  $X_{h''}$ ,  $h' < h''$ , resp., then either  $A \subseteq B$  or  $A \cap B = \emptyset$ . This gives rise to a tree structure representing the landscape [53, 32, 19, 24, 25, 39, 79], called the barrier tree. Local minima of the cost function correspond to the leaves of the tree, interior nodes are saddle points that define the fitness barriers between local minima. These notions of a local minimum and saddle point are clearly topological concepts, as they can be defined in terms of the behavior of the cost function in the neighborhood around a point:  $x$  is a local minimum if there is a neighborhood  $N$  of  $x$  such that  $f(y) \geq f(x)$  for all  $y \in N$ . The concept of a saddle point is much less clear. A number of similar but not-equivalent constructions is discussed in detail in [25]. A connection to combinatorial vector fields is drawn in [71].

**The Genotype-Phenotype Map.** Two distinct and largely independent processes drive biological evolution and lie at the heart of Charles Darwin’s theory: the generation of variation and selection of the variants according to their fitness. With the advent of molecular genetics it has become clear that variation is produced (primarily) at the molecular level by mutation, recombination, and other rearrangements of the genomic sequence information. Selection, in contrast, acts on the macroscopic living organism. The genotype-phenotype map, which relates genetic information to organismal appearance, properties, and behavior, thus plays a central role in modern theories of evolution [76].

RNA secondary structures have played a major role for the understanding of genotype-phenotype maps in general since they can be readily explored computationally [64]. In this setting, the primary sequence of the RNA is the genotype, while its secondary structure takes on the role of phenotype. At the phenotypic level, evolutionary processes are governed by the accessibility of phenotypic variants. The biophysical properties of RNA molecules allow at least a qualitative understanding of the likely structural

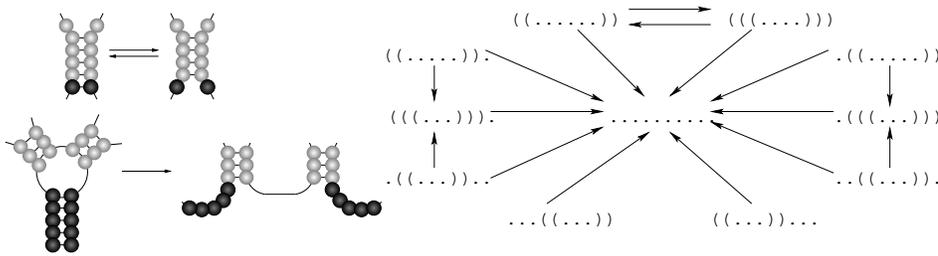


Figure 3: Fontana-Schuster topology of RNA phenotype space. L.h.s.: Frequent transitions between RNA secondary structures in response to point mutations are the opening or closing of single base pairs and the opening of constrained stems. The latter transition is not reversible since most sequences are not pre-disposed to closing a constrained stem in response to a point mutation. R.h.s.: Space of secondary structures from GC sequences of length 10. Arrows indicate accessibility defined as frequent transitions in the sense of the two rules on the r.h.s. Figure adapted from [12, 68].

effects of simple point mutations and thus the formulation of phenotypic rules of accessibility [28, 29], see Figure 3.

An evolutionary trajectory can be regarded as a function  $f$  from the time axis into phenotype space, where  $f(t)$  represents e.g. the dominating phenotype in a population at time  $t$ . Computer simulations of the RNA model reveal a pattern of periods of stasis with intermittent bursts of adaptive evolution. With few exceptions, consecutive phenotypes arise from each other through one of the easily accessible structural changes of Fig. 3. These transitions are indeed continuous in the usual topological sense [13, 73].

A phenotype is usually described by a set of “characters”, that is, properties that can be used to differentiate between different types of organisms. The particular state of a given character (e.g. the presence of 5 fingers) is interpreted as an evolutionary adaptation caused by natural selection. This explanation requires the assumption that the character state can be produced by mutation without significantly affecting the functionality and/or structure of the rest of the body, i.e., “quasi-independence” [57]. As argued in [80], this is a statement about the structure of accessible sets in phenotype space, namely the requirement that the phenotype space can be represented

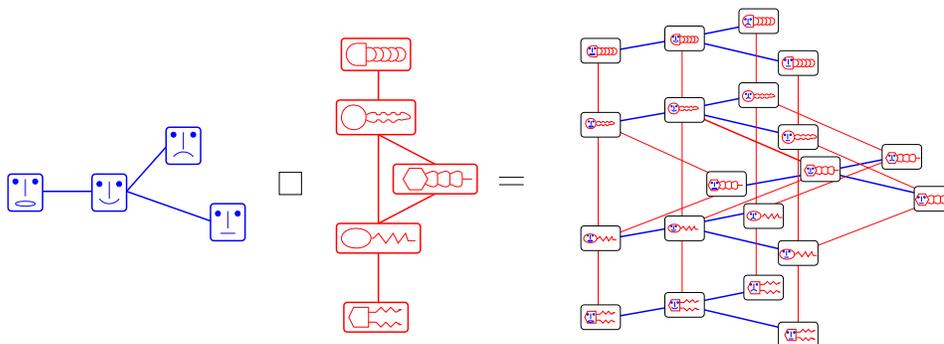


Figure 4: Phenotype space as a product of two characters (“body shape” and “faces”). For illustrative purpose the Cartesian graph product rather than the strong graph product, which corresponds to standard product of topological spaces, is shown. The latter is obtained by adding the two diagonal edges to each quadrangle. Figure adapted from [70].

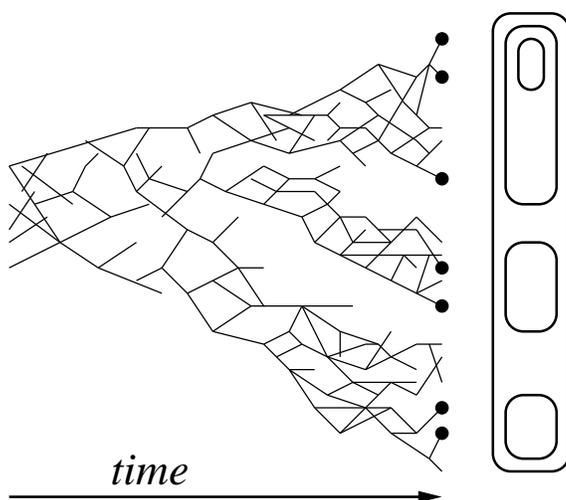


Figure 5: Sketch of the DAG  $G(V, E)$ . Fat dots indicate the vertices in the extant sample  $X \subset V$ . A possible clustering that conforms to the phylogenetic relationships is indicated. For details we refer to [20].

as a product of generalized topological space corresponding to the individual characters, Fig. 4. Since characters are meaningfully defined only in a local or regional subset of the entire phenotype space, the product structure also can only be a local approximation. These ideas have lead to the development of a theory of “Approximative Graph Products” and research into algorithms for local approximate factorization of graphs [47, 48].

**The “DAG of Everything”.** Instead of looking at the potential of evolutionary processes it is also of interest to consider only the actual history of life. This idea was recently explored in [20] using a graph  $G(V,A)$  whose vertex set  $V$  consists of all individuals that ever lived. A directed edge  $x \rightarrow y$  indicates that  $x$  contributed genetic information to  $y$ . Pedigree graphs are special cases of this construction, where the in-degree is limited to 2, namely an arc from the father and an arc from the mother, see e.g. [78, 63]. This model also easily incorporates horizontal gene transfer or the formation of hybrid species, and hence forms also a basis to study arbitrary complex phylogenetic networks [54]. Given a subset of extant observable individuals  $X \subset V$  several notions of connectedness of subsets of  $X$  can be defined. The main result of [20] is that these lead to natural collections of clusters that can be interpreted e.g. as phylogenies, Fig. 5. The DAG  $G(V,A)$  might also be an attractive starting point to study more general forms of phylogenetic networks.

## GENERALIZED TOPOLOGY

### Abstract closure functions

Instead of open sets, most approaches to generalized topology start from an abstract *closure function*  $c : \mathfrak{P}(X) \rightarrow \mathfrak{P}(X)$  that encapsulates a notion of reachability or accessibility from a given subset  $A$ . The natural conjugate  $i : \mathfrak{P}(X) \rightarrow \mathfrak{P}(X)$  defined by  $i(A) = X \setminus c(X \setminus A)$  identifies the interior of  $A$  as the part of  $A$  that is not accessible from the complement, i.e., the outside, of  $A$ . Closure and, equivalently, interior give rise to a concept of neighborhood of points:  $N$  is a neighborhood of  $x$ , if its interior contains  $x$ . The collection of neighborhoods can be viewed as a function  $\mathcal{N} : x \rightarrow \mathfrak{P}(\mathfrak{P}(X))$  so that  $N \in \mathcal{N}(x)$  iff  $x \in i(N)$ , i.e., iff  $x \notin c(X \setminus N)$ . Closure, interior, and neighborhood are equivalent in specifying the generalized topology [15]. For later reference we note that the notion of neighborhoods naturally extends to sets:  $N \in \mathcal{N}(A)$  iff  $A \subseteq i(N)$ , i.e.,  $N$  is a neighborhood of all  $x \in A$ .

The notion of continuity lies at the heart of topological theory. In the most abstract setting it comes in two flavors. Let  $(X, c_X)$  and  $(Y, c_Y)$  be two sets, each endowed with its closure function A function  $f : X \rightarrow Y$  is

*closure preserving* if for all  $A \in \mathcal{P}(X)$ ,  $f(c_X(A)) \subseteq c_Y(f(A))$  holds;  
*continuous* if for all  $B \in \mathcal{P}(Y)$ ,  $c_X(f^{-1}(B)) \subseteq f^{-1}(c_Y(B))$  holds.

One says that  $f : X \rightarrow Y$  is *continuous in  $x$*  if  $B \in \mathcal{N}(f(x))$  implies  $f^{-1}(B) \in \mathcal{N}(x)$ . It can be shown that  $f : X \rightarrow Y$  is continuous if and only if it is continuous in each  $x \in X$  [35, Thm.3.1.].

Obviously, the identity  $\iota : (X, c) \rightarrow (X, c) : x \mapsto x$  is both closure-preserving and continuous. Furthermore, the concatenation  $h = g \circ f$  of the closure-preserving (continuous) functions  $f : X \rightarrow Y$  and  $g : Y \rightarrow Z$  is again closure-preserving (continuous).

Note that at this point we have made no assumption at all on the properties of  $c$ . Almost all approaches to extend the framework of topology assume at least that the closure functions are isotonic [41, 42, 15, 7, 34]. The importance of isotony is emphasized by several equivalent conditions [42, Lem.10] listed in Tab. 1 below. A (not necessarily non-empty) collection  $\mathfrak{F} \subseteq \mathfrak{P}(X)$  is a *stack* if  $F \in \mathfrak{F}$  and  $F \subseteq G$  implies  $G \in \mathfrak{F}$ . The closure function  $c$  is isotonic if and only if the neighborhood system  $\mathcal{N}(x)$  is a stack for all  $x \in X$ . In isotone spaces, continuity and closure preservation are equivalent.

### Kuratowski's axioms

Kuratowski's axioms for the closure function of a topological space [55] may be seen as specializations of the very general closure functions that we have considered so far. It is interesting to note that each of them can be formulated equivalently for closure, interior, and neighborhoods. Different combinations of these axioms, summarized in Table 2 define generalized topological structures that have been studied to various degrees in the literature.

### Connectedness and Separation

Topological connectedness is closely related to separation. The basic idea is to investigate under which conditions closure or neighborhoods of distinct points or sets do or do not intersect. We say that  $A$  and  $B$  are *separated* if they have disjoint neighborhoods, i.e.,  $\mathcal{N}(A) \cap \mathcal{N}(B) = \emptyset$ . Two sets are *semi-separated* if there are neighborhoods  $N' \in \mathcal{N}(A)$  and  $N'' \in \mathcal{N}(B)$  such that  $A \cap N'' = N' \cap B = \emptyset$ . Consider a continuous function  $v : (X, c) \rightarrow$

Table 1: Kuratowski's axioms.

The properties below are meant to hold for all  $A, B \in \mathfrak{P}(X)$  and all  $x \in X$ , respectively.

	closure	interior	neighborhood
K0'	$\exists A : x \notin c(A)$	$\exists A : x \in i(A)$	$\mathcal{N}(x) \neq \emptyset$
K0	$c(\emptyset) = \emptyset$	$i(X) = X$	$X \in \mathcal{N}(x)$
K1 isotonic monotone	$A \subseteq B \implies c(A) \subseteq c(B)$ $c(A \cap B) \subseteq c(A) \cap c(B)$ $c(A) \cup c(B) \subseteq c(A \cup B)$	$A \subseteq B \implies i(A) \subseteq i(B)$ $i(A) \cup i(B) \subseteq i(A \cup B)$ $i(A \cap B) \subseteq i(A) \cap i(B)$	$N \in \mathcal{N}(x)$ and $N \subseteq N' \implies N' \in \mathcal{N}(x)$
KA	$c(X) = X$	$i(\emptyset) = \emptyset$	$\emptyset \notin \mathcal{N}(x)$
KB	$A \cup B = X \implies c(A) \cup c(B) = X$	$A \cap B = \emptyset \implies i(A) \cap i(B) = \emptyset$	$N', N'' \in \mathcal{N}(x) \implies N' \cap N'' \neq \emptyset$
K2 expansive	$A \subseteq c(A)$	$i(A) \subseteq A$	$N \in \mathcal{N}(x) \implies x \in N$
K3 sub-linear	$c(A \cup B) \subseteq c(A) \cup c(B)$	$i(A) \cap i(B) \subseteq i(A \cup B)$	$N', N'' \in \mathcal{N}(x) \implies N' \cap N'' \in \mathcal{N}(x)$
K4 idempotent	$c(c(A)) = c(A)$	$i(i(A)) = i(A)$	$N \in \mathcal{N}(x) \iff i(N) \in \mathcal{N}(x)$
K5 additive	$\bigcup_{i \in I} c(A_i) = c(\bigcup_{i \in I} A_i)$	$\bigcap_{i \in I} i(A_i) = i(\bigcap_{i \in I} A_i)$	$\mathcal{N}(x) = \emptyset$ or $\exists N(x) : N \in \mathcal{N}(x) \iff N(x) \subseteq N$

$[0, 1]$ , where  $[0, 1]$  denotes the unit interval endowed with the usual topology of the real numbers. It is called an *Urysohn function* separating  $A$  and  $B$  if  $v(A) \subseteq \{0\}$  and  $v(B) \subseteq \{1\}$ . If such a function exists,  $A$  and  $B$  are called Urysohn-separated.  $B$  is completely within  $A$ ,  $B \Subset A$ , if  $B$  and  $X \setminus A$  are Urysohn-separated. Urysohn-separated implies separated implies semi-separated.

A large number of subtly different *separation axioms* have been considered in the literature, of which here we just list a few to give the flavor. In the following, we consider conditions for all distinct points  $x, y$  and all disjoint non-empty subsets  $A, B$  in  $(X, c)$ .

(th0) there is  $z \in X$  and  $N \in \mathcal{N}(z)$  such that for all  $N' \subset N$ ,  $N' \in \mathcal{N}(z)$  hold  $x \in N'$  and  $y \notin N'$  or *vice versa*.

(th1)  $\mathcal{N}(x) = \mathcal{N}(y)$  implies  $x = y$ .

(T0) there is  $N \in \mathcal{N}(x)$  such that  $y \notin N$  or there is  $N' \in \mathcal{N}(y)$  such that

Table 2: Axioms for various types of closure space.

Defining axioms are marked by  $\bullet$ , further properties that are implied are marked by  $\circ$ .

Axiom	$c(\emptyset) = \emptyset$	$A \subseteq B \implies c(A) \subseteq c(B)$ isotonic	$A \subseteq c(A)$ enlarging	$c(A \cup B) \subseteq c(A) \cup c(B)$ sub-linear	$c(c(A)) = c(A)$ idempotent	$c(\bigcup_i A_i) = \bigcup_i c(A_i)$ additive	Ref.
	(K0)	(K1)	(K2)	(K3)	(K4)	(K5)	
Extended Topology	$\bullet$	$\bullet$					[42]
Brissaud	$\bullet$		$\bullet$				[7]
Neighborhood space	$\bullet$	$\bullet$	$\bullet$				[44]
Closure space	( $\bullet$ )	$\bullet$	$\bullet$		$\bullet$		[67]
Smyth space	$\bullet$	$\bullet$		$\bullet$			[66]
Binary relation	$\bullet$	$\circ$		$\circ$		$\bullet$	[58, 8]
Pretopology	$\bullet$	$\bullet$	$\bullet$	$\bullet$			[9]
Topology	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$		
Alexandroff space	$\bullet$	$\circ$	$\bullet$	$\circ$		$\bullet$	
Alexandroff topology	$\bullet$	$\circ$	$\bullet$	$\circ$	$\bullet$	$\bullet$	[2]

$$x \notin N'.$$

$$(T0') \quad y \notin c(\{x\}) \text{ or } x \notin c(\{y\}).$$

$$(T1) \quad \text{there is } N \in \mathcal{N}(x) \text{ and } N' \in \mathcal{N}(y) \text{ such that } x \notin N' \text{ and } y \notin N.$$

$$(T1') \quad c(x) \subseteq \{x\}.$$

$$(T2) \quad \text{there is } N \in \mathcal{N}(x) \text{ and } N' \in \mathcal{N}(y) \text{ such that } N \cap N' = \emptyset.$$

$$(T2\frac{1}{2}) \quad \text{there is } N \in \mathcal{N}(x) \text{ and } N' \in \mathcal{N}(y) \text{ such that } c(N) \cap c(N') = \emptyset.$$

(T2U) Any two distinct  $x$  and  $y$  are Urysohn-separated.

(R) If  $x \notin c(A)$  then there are disjoint neighborhoods of  $x$  and  $A$ .

(R') For every neighborhood  $N$  of  $x$  there is also a neighborhood  $U \in \mathcal{N}(x)$  such that  $c(U) \subseteq N$ .

(CR) For every neighborhood  $N$  of  $x$  there is also a neighborhood  $U \in \mathcal{N}(x)$  such that  $U \subseteq N$ .

(QN) If  $c(A) \cap c(B) = \emptyset$  then  $A$  and  $B$  are separated.

(N) If  $c(A) \cap c(B) = \emptyset$  then  $c(A)$  and  $c(B)$  are separated.

(UN) If  $c(A) \cap c(B) = \emptyset$  then  $A$  and  $B$  are Urysohn-separated.

(CN) Any two semi-separated sets are separated.

Several symmetry conditions are close associated with separation

(R0) If  $x$  is contained in every neighborhood of  $y$  then  $y$  is contained in every neighborhood of  $x$

(R0')  $x \in c(\{y\})$  implies  $y \in c(\{x\})$ .

(S) If  $x \in N$  for all  $N \in \mathcal{N}(y)$  then  $\mathcal{N}(y) = \mathcal{N}(x)$

(S') If  $c(A) \cap c(\{x\}) \neq \emptyset$  then  $x \in c(A)$ .

(RE) If  $N \cap N' \neq \emptyset$  for all  $N \in \mathcal{N}(x)$  and  $N' \in \mathcal{N}(x)$ , then  $\mathcal{N}(y) = \mathcal{N}(x)$ .

The interesting point about these axioms is that there are elaborate chains of implications among them in (ordinary) topological spaces. Some of these carry over to pretopologies, neighborhood spaces, or even isotonic spaces satisfying only (K0) and (K1), see Fig. 6 and its caption. Separation axioms have a close connections to generalized uniform structures [70] and the existence of metrics that allow the definition of  $\varepsilon$ -balls as basis for neighborhoods of points, see [9].

The separation axiom (T0) implies both *thinness* conditions (th0) and (th1), which do not seem to have been studied in detail in the context of topology. In particular (th1) plays an important role in the theory of graph products [49].

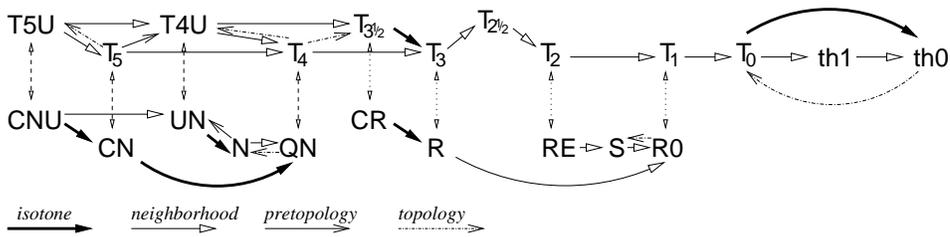


Figure 6: Relationships among separation and symmetry axioms. Different arrow types indicate implications that hold in all isotone space, in all neighborhood space, in all pretopological space, or only in topological spaces. The (T)-separation axioms are defined by a normality condition and (T1), dashed double-arrow, or by a regularity/symmetry condition and (T0), dotted double-arrow, respectively. Most of the proofs for the implications are non-trivial. They can be found in the book [9], in the supplemental material to [68], or in [69].

There have been several attempts to use connectedness as the primitive notion in topological theory [81, 43, 45]. Here, we employ the more usual approach to defined connectedness as a derived property in closure spaces: A set  $Z$  is *connected* in  $(X, c)$  if it is not a disjoint union of semi-separated pairs of non-empty sets  $A, Z \setminus A$ . In isotonic spaces, connected sets are characterized by the Hausdorff-Lennes condition:

$$[(A) \cap (Z \setminus A)] \cup [c(Z \setminus A) \cap A] \neq \emptyset \quad (4)$$

In neighborhood spaces, we arrive at a more familiar definition: A set is connected if and only if it is not the disjoint union of two non-empty closed (open) sets [38, Thm.5.2].

The collection of connected sets satisfies the following four properties in isotonic spaces [68, 38]:

- (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 and  $Z \subseteq c(Z)$ , then  $c(Z)$  is also connected.
- (c4) Let  $I$  be an arbitrary index set and  $x \in X$ . Suppose  $Z_i$  is connected and  $x \in Z_i$  for  $i \in I$ . Then  $W := \bigcup_{i \in I} Z_i$  is connected.

Consider a set  $A$  and a point  $x \in A$  and let  $A[x]$  be the union of all connected subsets of  $A$  that contain  $x$ . By (c4),  $A[x]$  is itself connected. We call it the connected component of  $A$  containing  $x$ . Note that the well-definedness of  $A[x]$  is required e.g. for the construction of the barrier trees mentioned in the Introduction.

The relationship of connected components and semi-separation becomes counter-intuitive in non-additive spaces: Suppose  $A$  has a finite number  $k > 1$  of connected components and let  $Q$  be such a component. Then  $Q$  and  $A \setminus Q$  are not necessarily semi-separated unless  $(X, c)$  is a pretopology [26]. Without the benefit of additivity, furthermore, it becomes natural to investigate alternative definitions of connectedness: In [5], for instance, a less stringent definition of connectivity is introduced that is in particular suitable for chemical reaction networks:  $A$  and  $B$  are *productively separated* if for all  $Z \subseteq A \cup B$  holds (1)  $c(Z \cap A) \cap B = \emptyset$  and  $c(Z \cap B) \cap A = \emptyset$ , and (2)  $c(Z) = c(Z \cap A) \cup c(Z \cap B)$ . A set is *productively connected* if it cannot be written as the disjoint union of two non-empty productively separated subsets.

Path-connectedness is a widely used notion of connectedness that in general is stronger than topological connectedness. From the topological point of view, a *path* is a *continuous* function  $p: [0, 1] \rightarrow X$  whose endpoints are  $p(0)$  and  $p(1)$ . A set  $A$  is *path-connected* if for any two points  $x, y \in A$ , there is a path  $p$  with  $p(0) = x$  and  $p(1) = y$ .

A useful lemma [26] characterizes connectedness in 2-point sets: Let  $(X, c)$  be a neighborhood space and  $\{x, y\} \subseteq X$  a 2-point subset. Then the following three statements are equivalent: (1)  $y \in c(\{x\})$  or  $x \in c(\{y\})$ ; (2)  $\{x, y\}$  is path-connected; (3)  $\{x, y\}$  is connected. We will return to this point later when considering finite generalized topologies.

## Product Spaces

Let  $(X_1, c_1)$  and  $(X_2, c_2)$  be two isotonic spaces. Then the product space  $(X_1 \times X_2, c_1 \times c_2)$  is defined by means of the neighborhood system  $\mathcal{N}(x_1, x_2)$ , where

$$N \in \mathcal{N}(x_1, x_2) \iff \exists N_1 \in \mathcal{N}_1(x_1) \text{ and } N_2 \in \mathcal{N}_2(x_2) \text{ such that } N_1 \times N_2 \subseteq N \quad (5)$$

For sets of the form  $A_1 \times A_2$  this translates to  $c(A_1 \times A_2) = c_1(A_1) \times c_2(A_2)$  in the product space, see [34, Thm.8.1]. If  $(X_1, c_1)$  and  $(X_2, c_2)$  satisfy

(K2), (K3), or (K4), respectively, then so does their product. We remark that the construction of products can be extended to infinite families of spaces, even uncountable ones. The projections  $\pi_i : \prod_j (X_j, c_j) \rightarrow (X_i, c_i) : x = (\dots, x_i, \dots) \mapsto x_i$  are continuous.

The *inductive product*  $(X_1, c_1) \square (X_2, c_2)$  has the neighborhoods  $N \in \mathcal{N}^\square(x_1, x_2)$  iff there is  $N_1 \in \mathcal{N}_1(x_1)$  and  $N_2 \in \mathcal{N}_2(x_2)$  such that  $N_1 \times \{x_2\} \cup \{x_1\} \times N_2 \subseteq N$ . This product is discussed briefly in [9].

## Finite Generalized Topologies

For practical applications in computational biology and computational chemistry, finite spaces are of particular interest. Of course, in this case (K3) implies (K5). The by far best studied finite structure are the finite pretopologies: these are the simple directed graphs. More precisely, if  $(X, c)$  is a finite pretopological space, the associated graph has vertex set  $X$  and  $(x, y)$  is a directed edge iff  $y \in c(\{x\})$ . Conversely, given a graph this recipe defines a finite pretopology on  $X$  [73]. Graphs that correspond to topological spaces are considered e.g. in [13].

Several of the topological concepts outlined above have independently have been developed in graph theory. Continuous function, for instance, coincide with graph homomorphisms [46], i.e., functions  $f : (X, E) \rightarrow (Y, F)$  such that  $(x, y) \in E$  implies  $(f(x), f(y)) \in F$  or  $f(x) = f(y)$ . The usual graph-theoretical definition of connectedness is also the same as pre-topological connectedness: a two-element subset  $\{x, y\}$  is connected if and only if  $(x, y) \in E$  or  $(y, x) \in E$ . Furthermore, connectedness and path-connectedness is the same [26]. The (R0) symmetry axiom, furthermore, characterizes undirected graphs. On the other hand, it appears to be unknown if strong connectedness in digraphs has a straightforward interpretation as a topological property.

The strong product of graphs coincides with the product of generalized topological spaces, while the inductive product is the same as the Cartesian graph product. For graphs, products have been studied in particular w.r.t. to the conditions under which a graph has a unique prime factorization. We refer to the book [40] for an extensive discussion of this topic. Interestingly, thinness conditions play an important role in this context. Local and approximative product structures have been explored recently [47, 48] motivated by the interpretation of characters as local factor spaces of phe-

notype space [80].

As mentioned in the introduction, topological spaces are typically studied in terms of their closed sets, i.e., the collection  $\mathcal{C} = \{A | c(A) = A\}$ . If (K1) to (K4) are satisfied, we have (I1)  $X \in \mathcal{C}$ , (I2) arbitrary intersections of closed sets are closed, and (I3) the union of two closed sets is closed. In lattice theory more general so-called *intersection structures* are considered that fulfill only (I2) see e.g., [14], in graph theory the same structures appear as convexities [11]. Since neither idempotency nor additivity of the closure function readily applies to the examples in the introduction we will not consider them further in this contribution.

Not much is known about finite neighborhood spaces. They correspond to the  $\mathfrak{N}$ -systems introduced in [50]. Consider a pair  $(X, \mathfrak{N})$  consisting of a nonempty finite set  $X$  and a function  $\mathfrak{N} : X \rightarrow \mathcal{P}(\mathcal{P}(X))$  that associates to each  $x \in X$  a collection  $\mathfrak{N}(x) = \{N^1(x), N^2(x), \dots, N^{d(x)}\}$  of  $d(x)$  subsets of  $X$  with the following properties:

(N0)  $\mathfrak{N}(x) \neq \emptyset$ .

(N1)  $N^i(x) \subseteq N^j(x)$  implies  $i = j$ .

(N2)  $x \in N^i(x)$ , for  $1 \leq i \leq d(x)$ .

$\mathfrak{N}$ -systems are by construction exactly the finite neighborhood spaces, when the  $N^i(x)$  are interpreted as minimal neighborhoods, i.e.,  $N \in \mathcal{N}(x)$  if and only if there is  $N' \subseteq N$  with  $N' \in \mathfrak{N}(x)$ . We may also interpret them as a special type of directed hypergraphs with hyperedges of the form  $(\{x\}, N^i(x))$ . Axiom (N1) ensures that it is simple.

The characterization of connectedness in 2-point sets [26] suggests to associate a graph  $\vec{\Gamma}(X, c)$  with vertex set  $X$  with a given finite neighborhood space  $(X, c)$  such that  $(x, y)$  is a (directed) edge if  $y \in c(\{x\})$ . If  $c$  is additive, then  $\vec{\Gamma}$  is exactly the graph representation of the pretopological space  $(X, c)$ . In finite neighborhood spaces every path consists of a finite number of connected 2-point sets consisting of consecutive points. Thus a subset of  $X$  is path-connected if and only if the corresponding induced subgraph of  $\vec{\Gamma}$  is path connected [26]. Connectedness is strictly weaker property since a connected set with three points does not necessarily contain connected pairs. Finite neighborhood spaces admit a unique prime factorization w.r.t. to the usual topological product under certain conditions on  $\vec{\Gamma}$  [50].

Finite isotone spaces, finally, have not been considered as combinatorial objects to our knowledge.

## DIRECTED HYPERGRAPHS AS TOPOLOGICAL STRUCTURES

It may come as a surprise that hypergraphs have remained virtually unstudied from a topological perspective. While the interpretation of graphs as finite pretopological spaces is quite natural, we will see below that there does not seem to be a unique canonical translation for hypergraphs.

A directed hypergraph  $H$  consists of a vertex set  $X$  and a set  $\mathcal{E}$  of directed hyperedges, each of which is a pair  $E = (E^-, E^+)$  with  $E^-, E^+ \subseteq X$ . A hypergraph is simple if no edge is properly contained in another one, i.e., if  $E^- \subseteq F^-$  and  $E^+ \subseteq F^+$  implies  $E = F$ . One way of defining undirected hypergraphs as a special case of directed ones is to require  $E^+ = E^-$  for all  $E \in \mathcal{E}$ .

Given a directed hypergraph  $H$ , it appears natural to consider the closure function

$$c(A) := \bigcup_{\mathcal{E}: E^- \subseteq A} E^+ \quad (6)$$

or

$$c(A) := A \cup \bigcup_{\mathcal{E}: E^- \subseteq A} E^+ \quad (7)$$

depending on whether we want to insist that  $c$  is enlarging or not. This approach indeed has been taken in [5] to describe chemical reaction networks. These constructions, however, become trivial for undirected hypergraphs since the condition  $E^+ = E^-$  for all hyperedges ensures that  $c(A) \subseteq A$ .

As a possible remedy one might identify undirected hypergraphs with directed ones that satisfy the following symmetry condition:

(SH) For every directed hyperedge  $\mathcal{E} = (E^-, E^+)$  every pair  $(A, B)$  of non-empty sets with  $A \cup B = E^- \cup E^+$  is also a hyperedge.

In particular, in this case  $(\{x\}, E)$  with  $E := E^- \cup E^+$  and any  $x \in E$  is a hyper-edge. Thus, for a given hyperedge  $\mathcal{E}$ , if  $E^- \subseteq A$  then there  $x \in E \cap A$  and hence  $(\{x\}, E)$  also contributes to  $c(A)$ . Thus equ.(6) can be rephrased as

$$c(A) = \bigcup_{\mathcal{E}: E \cap A \neq \emptyset} E \quad (8)$$

This expression also makes perfect sense for undirected hypergraphs. It is equally unsatisfactory, however, since this closure function is obviously additive and hence describes a (symmetric) graph with adjacency relation  $x \sim y$  iff there a hyperedge  $\mathcal{E}$  such that  $x, y \in E$ .

An alternative approach is to consider hyperedges as minimal neighborhoods in the sense of the  $\mathfrak{N}$ -systems introduced in the previous section. In finite pretopologies, i.e. directed graphs, the intersection of all  $N \in \mathcal{N}(x)$  is again a neighborhood of  $x$ . It coincides with the “in-neighborhood” of  $x$ , given by

$$N(x) = \{y \in X \mid x \in c(y)\} \quad (9)$$

For directed hypergraphs, one would analogously postulate that hyper-edges that “produce”  $x$  are the minimal neighborhoods of  $x$ , i.e.,

$$N \in \mathcal{N}(x) \iff \exists \mathcal{E} : x \in E^+ \text{ and } E^- \cup E^+ \subseteq N \quad (10)$$

It is straightforward to compute the corresponding closure function

$$c(A) = \{x \mid \exists \mathcal{E} : x \in E^+ \text{ implies } (E^- \cup E^+) \cap A \neq \emptyset\} \quad (11)$$

An undesirable feature remains, however. If we specialize this construction to directed graphs, i.e., so that  $E^-$  and  $E^+$  consist of two distinct points for each hyperedge, then  $N \in \mathcal{N}(x)$  iff there is an arc  $(y, x)$  such that  $x, y \in N$ . The associated closure function is  $c(A) = A \cup \{z \mid N_{in}(z) \subseteq A\}$ . This is in general not additive.

A possibly fruitful alternative is to start from the notation of hypergraph homomorphisms  $(X, \mathcal{E}) \rightarrow (Y, \mathit{mathcal{F}})$ , i.e., maps  $\varphi : X \rightarrow X'$  so that for every  $(E^-, E^+) \in \mathcal{E}$  there is  $(F^-, F^+) \in \mathcal{F}$  so that  $\varphi(E^-) \subseteq F^-$  and  $\varphi(E^+) \subseteq F^+$ . One may then ask if there exist isotonic closure functions so that the hypergraph homomorphisms are exactly the continuous function as in the case of graphs. To our knowledge, this question is open.

We have to conclude that at present there is no satisfactory interpretation of hypergraphs a topological objects.

## DYNAMICAL ASPECTS

The notion of *topological dynamical systems*, see e.g. [16] does not seem to have explored in systematic way for generalized topological spaces. A natural starting point would be flows of the form  $\omega : T \times X \rightarrow X$  where  $T$

is suitable (topological) semigroup such as the non-negative integers and  $(\iota, x) \mapsto x$ , where  $\iota$  denotes the neutral element of  $T$ . Orbits, for instance, are well-defined in this setting:  $\Omega(x) = \{tx | t \in T\} \subseteq X$ , where  $tx$  denotes the action of the semigroup element  $t$  on the point  $x$ . Similarly, trajectories are simply functions of the form  $\xi : T \rightarrow X$  such that  $t \mapsto \xi(t) = tx_0$  for a fixed “initial condition”  $x_0$ . In [73], for instance, conditions are explored under which trajectories of phenotypic evolution are continuous. It turns out that this matches well with intuition developed earlier in [28, 29] in the context of the evolution of RNA secondary structures.

Combinatorial vector fields offer a more specialized alternative formalization. Let  $M$  be a simplicial complex constructed over the set  $K$  of simplices, see e.g. [51]. We write  $\sigma < \tau$  if the simplex  $\sigma$  lies in the boundary of the simplex  $\tau$ . A *combinatorial vector field* on  $M$  [30] is a map  $\eta : K \rightarrow K \cup \{\emptyset\}$  such that

1. If  $\eta(\sigma) \neq \emptyset$  then  $\dim \eta(\sigma) = \dim(\sigma) + 1$ , and  $\sigma < \eta(\sigma)$ .
2. If  $\eta(\sigma) = \tau \neq \emptyset$ , then  $\eta(\tau) = \emptyset$ .
3. For all  $\sigma \in K$ ,  $|\sigma^{-1}| \leq 1$ .

We remark that combinatorial vector fields can in fact be defined on the much more general CW-complexes [30].

Combinatorial vector fields come with a natural notion of  $\eta$ -paths as the analog of trajectories, namely a finite sequence of simplices  $\gamma = (\sigma_0, \tau_0, \sigma_1, \tau_1, \dots, \sigma_{n-1}, \tau_{n-1}, \sigma_n)$  such that  $\eta(\sigma_i) = \tau_i$  for  $0 \leq i < n$  and  $\sigma_{i+1} < \tau_i$ . A rest point is a simplex  $\sigma$  such that  $\eta(\sigma) = \emptyset$  and  $\eta^{-1}(\sigma) = \emptyset$ . The rest points and the closed  $\eta$ -paths together form the so-called chain-recurrent set  $\mathcal{R}$  of  $\eta$ . These play the role of attractors. All other trajectories lead towards  $\mathcal{R}$ .

A *Lyapunov function* of the combinatorial vector field  $\eta$  is a function  $F : M \rightarrow \mathbb{R}$  such that

1. if  $\sigma \notin \mathcal{R}$  and  $\tau > \sigma$  then
  - (a)  $F(\sigma) < F(\tau)$  if  $\tau \neq \eta(\sigma)$
  - (b)  $F(\sigma) \geq F(\tau)$  if  $\tau = \eta(\sigma)$
2. if  $\sigma \in \mathcal{R}$  and  $\tau > \sigma$  then

- (a)  $F(\sigma) = F(\tau)$  if  $\sigma \sim \tau$
- (b)  $F(\sigma) < F(\tau)$  if  $\sigma \not\sim \tau$

A combinatorial version of Conley's theorem ensures that there is Lyapunov function for every combinatorial vector field [30]. Combinatorial vector fields thus form discrete analog for gradient vector fields.

In [71] this observation was used to consider a the collection of combinatorial vector fields on a graph  $G$  (with vertex set  $X$ ) for which a given function  $F : X \rightarrow \mathbb{R}$  is a Lyapunov function. In this context, the function  $F$  is interpreted as energy function. The combinatorial vector field, on the other hand, can be interpreted as a partial orientation  $P$  on  $G$  so that  $\{x, y\}$  is directed from  $x$  to  $y$  iff  $\eta(x) = \{x, y\}$ . Conversely a system of directed edges  $P$  with the properties

1.  $(x, y) \in P$  implies  $\{x, y\} \in E$  (consistency with  $G$ )
2.  $(x, y) \in P$  and  $(x, z) \in P$  implies  $y = z$  (uniqueness)
3.  $(x, y) \in P$  implies  $(y, x) \notin P$  (antisymmetry)

corresponds to a combinatorial vector field on  $G$ . A trajectory in  $\eta$  is then given by a sequence of vertices  $(x_i), i = 1, \dots, k$  so that  $(x_i, x_{i+1}) \in P$ .

These combinatorial vector fields provide a convenient description of the system of adaptive (downhill) walks on the energy landscape  $(G, F)$  and admits an alternative approach towards characterizing basins, barriers, and their hierarchical structure. In particular, they highlight the complications in the analysis of landscapes arising from degeneracies in the energy function. Given  $F$ , denote by  $G^F$  the subgraph of  $G$  with edges  $\{x, y\} \in E(G)$  if  $F(x) = F(y)$ . The connected components  $G^F(x)$ , the so called shelves of the landscape, provide the main complication for practical computations, Fig. 7. While the vector fields for which  $F$  is a Lyapunov function necessarily point downwards between shelves, it is complicated to handle the possible orientations with a single shelf since edges may be oriented in both directions depending on the particular choice of  $\eta$ .

A direct connection to the topology of the landscape  $(G, F)$  is obtained by mean of the following construction:

**Definition.** A point  $y$  is reachable from  $x$  in  $(G, F)$  if there is a combinatorial vector field  $\eta$  for which  $F$  is a Lyapunov function and that admits a trajectory from  $x$  to  $y$ . Let  $C(x)$  be the set of vertices reachable from  $x$ .

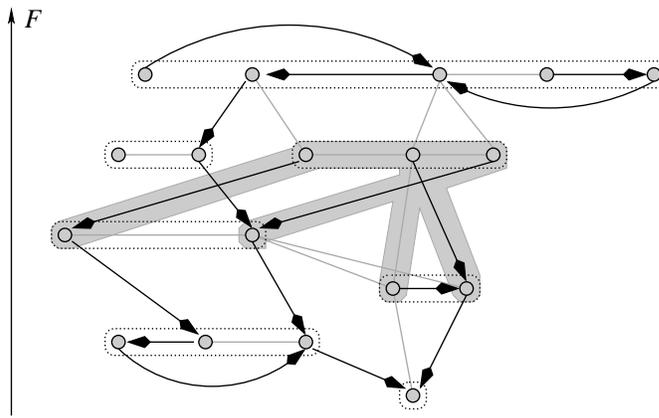


Figure 7: Shelves, i.e., maximal connected sets on which the energy function  $F$  is constant are indicated by dotted boxes. Edges of  $G$  that are oriented by the combinatorial vector field  $\eta$  are shown as arrows, the remaining edges of  $G$  are shown in gray. Edges between shelves are either not oriented or point downwards in any combinatorial vector field for which  $F$  is a Lyapunov function. Both orientations are possible for edges within shelves. Figure adapted from [71].

As shown in [71],  $C$  can be extended to an additive closure function that satisfies all of Kuratowski’s axioms and hence defined a finite topological space on the vertex set  $X$ . This topology is intimately related with the barrier structure of the landscape by means of the notion of a valley: A *valley* in  $(G, F)$  is a maximal connected closed set with respect to the reachability topology. Equivalently a valley  $W$  is a maximal connected subset of  $X$  so that no vertex  $y \notin W$  is reachable from any  $x \in W$ .

## CONCLUDING REMARKS: COARSE GRAINING

Accessibility is of particular interest in the context of “constructive systems”, i.e., models comprising a set combinatorial objects and set of rules with which they can be combined. This was pioneered e.g. in the work of Fontana and Buss [27] using  $\lambda$ -calculus as a modelling platform for abstract chemistry. The state space  $X$  in such models is usually not finite. In simulations of (evolutionary) processes on such models, families of ob-

jects emerge that share many regularities and differ essentially in size only [27]. In chemistry, polymers are the most obvious example. Naturally, the question arises whether there are coarse-grained representations.

In the analysis of the fine structure of the Network of Organic Chemistry, it has turned out that so-called one-pot reactions, i.e., reactions that can be performed concurrently without the need to purify intermediates and hence are of practical interest in practice, play an important role [36]. Since one-pot reaction lumps together individual reactions into sequences they can be seen as a form of coarse-graining of the chemical space. Recently, we have started to explore [3], from an algebraic perspective, whether compositions of transformation rules can be employed to capture families such as homologous series (i.e., groups of chemicals with analogous properties that differ only in the number of simple structural units) based on their reactive capabilities.

Coarse-graining is an important issue also in the analysis e.g. of fitness landscapes. A rather simple example is the use of barrier trees, which represent a partition of state space  $X$ . Quotient spaces  $X/\sim$ , obtained by identifying  $\sim$ -equivalent points of  $X$ , thus appear to be a natural formal structure to consider in this context. To our knowledge, this avenue has not been explored systematically for any of the model systems discussed above. It will be interesting to see if the intuitive connection of rule compositions and quotient spaces can be given a precise topological meaning.

### Acknowledgments

This work was supported in part by the Volkswagen Stiftung and the *Deutsche Forschungsgemeinschaft* within the EUROCORES Programme EUROGIGA (project GReGAS) of the European Science Foundation.

### References

- [1] S. A. Adeleke and P. M. Neumann. *Relations Related to Betweenness: Their Structure and Automorphisms*. Amer. Mathematical Society, 1998.
- [2] P. Alexandroff. Diskrete Räume. *Math. Sb. (N.S.)*, 2:501–518, 1937.

- [3] J. L. Andersen, C. Flamm, D. Merkle, and P. F. Stadler. Inferring chemical reaction patterns using graph grammar rule composition. 2012. submitted.
- [4] J. Behre, L. F. de Figueiredo, S. Schuster, and C. Kaleta. Detecting structural invariants in biological reaction networks. *Methods Mol Biol.*, 804:377–407, 2012.
- [5] G. Benkő, F. Centler, P. Dittrich, C. Flamm, B. M. R. Stadler, and P. F. Stadler. A topological approach to chemical organizations. *Alife*, 15:71–88, 2009.
- [6] K. J. M. Bishop, R. Klajn, and B. A. Grzybowski. The core and most useful molecules in organic chemistry. *Angew. Chem. Int. Ed.*, 45:5348–5354, 2006.
- [7] M. M. Brissaud. Les espaces prétopologiques. *C. R. Acad. Sc. Paris Ser. A*, 280:705–708, 1975.
- [8] C. Calude and V. E. Căzănescu. On topologies generated by Mosil resemblance relations. *Discr. Math.*, 25:109–115, 1979.
- [9] E. Čech. *Topological Spaces*. Wiley, London, 1966.
- [10] M. Changat, S. Klavžar, and H. M. Mulder. The all-path transit function of a graph. *Czech. Math. J.*, 51:439–448, 2001.
- [11] M. Changat, S. Klavžar, H. M. Mulder, and A. Vijayakumar. *Convexity in discrete structures*. Ramanujan Mathematical Society Lecture Notes Series, Vol. 5. International Press of Boston, 2010.
- [12] J. Cupal, I. L. Hofacker, and P. F. Stadler. Dynamic programming algorithm for the density of states of RNA secondary structures. In R. Hofstädt, T. Lengauer, M. Löffler, and D. Schomburg, editors, *Computer Science and Biology 96 (Proceedings of the German Conference on Bioinformatics)*, pages 184–186, Leipzig, Germany, 1996. Universität Leipzig.
- [13] J. Cupal, S. Kopp, and P. F. Stadler. RNA shape space topology. *Alife*, 6:3–23, 2000.

- [14] B. A. Davey and H. A. Priestley. *Introduction to Lattice and Order*. Cambridge Univ. Press, Cambridge UK, 1990.
- [15] M. M. Day. Convergence, closure, and neighborhoods. *Duke Math. J.*, 11:181–199, 1944.
- [16] J. de Vries. *Elements of topological dynamics*. Kluwer, Dordrecht, NL, 2010.
- [17] K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. Wiley, Chichester, NY, 2001.
- [18] P. Dittrich and P. Speroni di Fenizio. Chemical organization theory. *Bull. Math. Biol.*, 69:1199–1231, 2007.
- [19] J. P. Doye, M. A. Miller, and D. J. Welsh. Evolution of the potential energy surface with size for Lennard-Jones clusters. *J. Chem. Phys.*, 111:8417–8429, 1999.
- [20] A. W. M. Dress, V. Moulton, M. Steel, and T. Wu. Species, clusters and the “tree of life”: A graph-theoretic perspective. *J. Theor. Biol.*, 265:535–542, 2010.
- [21] U. Eckhardt and L. Latecki. Digital topology. Technical Report 89, Hamburger Beitr. z. Angew. Math. A, 1994.
- [22] H. E. F. LeBourgeois, M. Bouayad. Structure relation between classes for supervised learning using pretopology. In *Fifth International Conference on Document Analysis and Recognition*, page 33, 1999.
- [23] M. Fialkowski, K. J. M. Bishop, V. A. Chubukov, C. J. Campbell, and B. A. Grzybowski. Architecture and evolution of organic chemistry. *Angew. Chem. Int. Ed.*, 44:7263–7269, 2005.
- [24] C. Flamm, W. Fontana, I. Hofacker, and P. Schuster. RNA folding kinetics at elementary step resolution. *RNA*, 6:325–338, 2000.
- [25] C. Flamm, I. L. Hofacker, P. F. Stadler, and M. T. Wolfinger. Barrier trees of degenerate landscapes. *Z. Phys. Chem.*, 216:155–173, 2002.

- [26] C. Flamm, B. M. R. Stadler, and P. F. Stadler. Saddles and barrier in landscapes of generalized search operators. In C. R. Stephens, M. Toussaint, D. Whitley, and P. F. Stadler, editors, *Foundations of Genetic Algorithms IX*, volume 4436 of *Lecture Notes Comp. Sci.*, pages 194–212, Berlin, Heidelberg, 2007. Springer. 9th International Workshop, FOGA 2007, Mexico City, Mexico, January 8–11, 2007.
- [27] W. Fontana and L. W. Buss. What would be conserved if ‘the tape were played twice’? *Proc Natl Acad Sci USA*, 91:757–761, 1994.
- [28] W. Fontana and P. Schuster. Continuity in Evolution: On the Nature of Transitions. *Science*, 280:1451–1455, 1998.
- [29] W. Fontana and P. Schuster. Shaping Space: The Possible and the Attainable in RNA Genotype-Phenotype Mapping. *J. Theor. Biol.*, 194:491–515, 1998.
- [30] R. Forman. Combinatorial vector fields and dynamical systems. *Math. Z.*, 228:629–681, 1998.
- [31] A. Galton. Continuous motion in discrete space. In A. G. Cohn, F. Giunchiglia, and B. Selman, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Seventh International Conference (KR2000)*, pages 26–37, San Francisco, CA, 2000. Morgan Kaufmann Publishers.
- [32] P. Garstecki, T. X. Hoang, and M. Cieplak. Energy landscapes, supergraphs, and “folding funnels” in spin systems. *Phys. Rev. E*, 60:3219–3226, 1999.
- [33] P. Gitchoff and G. P. Wagner. Recombination induced hypergraphs: a new approach to mutation-recombination isomorphism. *Complexity*, 2:37–43, 1996.
- [34] S. Gnińska. On extended topologies. I: Closure operators. *Ann. Soc. Math. Pol., Ser. I, Commentat. Math.*, 34:81–94, 1994.
- [35] S. Gnińska. On continuity in extended topologies. *Ann. Soc. Math. Pol., Ser. I, Commentat. Math.*, 37:99–108, 1997.

- [36] C. M. Gothard, S. Soh, N. A. Gothard, B. Kowalczyk, Y. Wei, B. Baytekin, and B. A. Grzybowski. Rewiring chemistry: Algorithmic discovery and experimental validation of one-pot reactions in the network of organic chemistry. *Angew. Chem. Int. Ed.*, 51:7922–7927, 2012.
- [37] B. A. Grzybowski, K. J. M. Bishop, B. Kowalczyk, and C. E. Wilmer. The 'wired' universe of organic chemistry. *Nature Chemistry*, 1:31–36, 2009.
- [38] E. D. Habil and K. A. Elzenati. Connectedness in isotonic spaces. *Turk. J. Math*, 30:247–262, 2006.
- [39] J. Hallam and A. Prügel-Bennett. Large barrier trees for studying search. *IEEE Trans. Evol. Comput.*, 9:385–397, 2005.
- [40] R. Hammack, W. Imrich, and S. Klavžar. *Handbook of Product graphs*. CRC Press, Boca Raton, 2nd edition, 2011.
- [41] P. C. Hammer. General topology, symmetry, and convexity. *Trans. Wisconsin Acad. Sci., Arts, Letters*, 44:221–255, 1955.
- [42] P. C. Hammer. Extended topology: Set-valued set functions. *Nieuw Arch. Wisk. III*, 10:55–77, 1962.
- [43] P. C. Hammer. Extended topology: Connected sets and Wallace separations. *Portug. Math.*, 22:77–93, 1963.
- [44] P. C. Hammer. Extended topology: Continuity I. *Portug. Math.*, 25:77–93, 1964.
- [45] J. M. Harris. Continuity and separation for point-wise symmetric isotonic closure functions. Technical report, 2005. arXiv:math.GN/0507230.
- [46] P. Hell and Nešetřil. *Graphs and homomorphisms*. Oxford University Press, Oxford, UK, 2004.
- [47] M. Hellmuth, W. Imrich, W. Klöckl, and P. F. Stadler. Approximate graph products. *Eur. J. Comb.*, 30:1119–1133, 2009.

- [48] M. Hellmuth, W. Imrich, W. Klöckl, and P. F. Stadler. Local algorithms for the prime factorization of strong product graphs. *Math. Comp. Sci.*, 2:653–682, 2009.
- [49] W. Imrich and S. Klavžar. *Product Graphs: Structure and Recognition*. Wiley, New York, 2000.
- [50] W. Imrich and P. F. Stadler. A prime factor theorem for a generalized direct product. *Discussiones Math. Graph Th.*, 26:135–140, 2006.
- [51] J. Jonsson. *Simplicial Complexes of Graphs*. Springer, Berlin, 2007.
- [52] P. D. Karp and R. Caspi. A survey of metabolic databases emphasizing the MetaCyc family. *Arch Toxicol.*, 85:1015–1033, 2011.
- [53] T. Klotz and S. Kobe. “Valley Structures” in the phase space of a finite 3D Ising spin glass with  $\pm i$  interactions. *J. Phys. A: Math. Gen.*, 27:L95–L100, 1994.
- [54] V. Kunin, L. Ogldovsky, N. Darzentas, and A. Ouzounis. The net of life: reconstructing the microbial phylogenetic network. *Genome Res.*, 15:954–959, 2005.
- [55] C. Kuratowski. Sur la notion de limite topologique d’ensembles. *Ann. Soc. Polon. Math.*, 21:219–225, 1949.
- [56] C. LARGERON and S. Bonnevey. A pretopological approach for structural analysis. *Information Sciences*, 144:169–185, 2002.
- [57] R. C. Lewontin. Adaptation. *Sci. Am.*, 239:156–169, 1978.
- [58] M. Malitza. Topology, binary relations, and internal operations. *Rev. Roumaine Math. Pures Appl.*, 4:515–519, 1975.
- [59] S. Marchand-Maillet and Y. M. Sharaiha. Discrete convexity, straightness, and the 16-neighborhood. *Computer Vision & Image Underst.*, 66:316–329, 1997.
- [60] E. Pełkalska and R. P. W. Duin. *The Dissimilarity Representation for Pattern Recognition: Foundations and Applications*. World Scientific, Singapore, 2005.

- [61] S. Peter and P. Dittrich. On the relation between organizations and limit sets in chemical reaction systems. *Adv. Complex Syst.*, 14:77–96, 2011.
- [62] J. Pfaltz. Closure lattices. *Discrete Mathematics*, 154:217–236, 1996.
- [63] M. Riemer, P. F. Stadler, and K. Klemm. FRANz: Fast reconstruction of wild pedigrees. *Bioinformatics*, 25:2134–2139, 2009.
- [64] P. Schuster, W. Fontana, P. F. Stadler, and I. L. Hofacker. From sequences to shapes and back: A case study in RNA secondary structures. *Proc. Roy. Soc. Lond. B*, 255:279–284, 1994.
- [65] M. Shpak and G. P. Wagner. Asymmetry of configuration space induced by unequal crossover: implications for a mathematical theory of evolutionary innovation. *Artificial Life*, 6:25–43, 2000.
- [66] M. B. Smyth. Semi-metric, closure spaces and digital topology. *Theor. Computer Sci.*, 151:257–276, 1995.
- [67] W. P. Soltan. *An Introduction in Axiomatic Theory of Convexity*. Shtiintsa, Kishinev, 1984. Russian.
- [68] B. M. R. Stadler and P. F. Stadler. Generalized topological spaces in evolutionary theory and combinatorial chemistry. *J. Chem. Inf. Comput. Sci.*, 42:577–585, 2002. Proceedings MCC 2001, Dubrovnik.
- [69] B. M. R. Stadler and P. F. Stadler. Higher separation axioms in generalized closure spaces. *Commentationes Math. Warszawa, Ser. I*, 43:257–273, 2003.
- [70] B. M. R. Stadler and P. F. Stadler. The topology of evolutionary biology. In Ciobanu, editor, *Modeling in Molecular Biology*, Natural Computing Series, 267–286, 2004. Springer Verlag.
- [71] B. M. R. Stadler and P. F. Stadler. Combinatorial vector fields and the valley structure of fitness landscapes. *J. Math. Biol.*, 61:877–898, 2010.
- [72] B. M. R. Stadler, P. F. Stadler, M. Shpak, and G. P. Wagner. Recombination spaces, metrics, and pretopologies. *Z. Phys. Chem.*, 216:217–234, 2002.

- [73] B. M. R. Stadler, P. F. Stadler, G. P. Wagner, and W. Fontana. The topology of the possible: Formal spaces underlying patterns of evolutionary change. *J. Theor. Biol.*, 213:241–274, 2001.
- [74] P. F. Stadler and C. Flamm. Barrier trees on poset-valued landscapes. *Genetic Prog. Evol. Mach.*, 7-20:4, 2003.
- [75] P. F. Stadler, R. Seitz, and G. P. Wagner. Evolvability of complex characters: Population dependent Fourier decomposition of fitness landscapes over recombination spaces. *Bull. Math. Biol.*, 62:399–428, 2000.
- [76] P. F. Stadler and B. M. R. Stadler. Genotype phenotype maps. *Biological Theory*, 3:268–279, 2006. Konrad Lorenz Institute Workshop on Biological Information organized by Werner Callebaut in 2002.
- [77] P. F. Stadler and G. P. Wagner. The algebraic theory of recombination spaces. *Evol. Comp.*, 5:241–275, 1998.
- [78] M. Steel and J. Hein. Reconstructing pedigrees: A combinatorial perspective. *J. Theor. Biol.*, 240:360–367, 2006.
- [79] E. Tøstesen. Partly melted DNA conformations obtained with a probability peak finding method. *Phys. Rev. E*, 71:061922, 2005.
- [80] G. Wagner and P. F. Stadler. Quasi-independence, homology and the unity of type: A topological theory of characters. *J. Theor. Biol.*, 220:505–527, 2003.
- [81] A. D. Wallace. Separation spaces. *Ann. Math.*, pages 687–697, 1941.
- [82] A. V. Zeigarnik. On hypercycles and hypercircuits in hypergraphs. In P. Hansen, P. W. Fowler, and M. Zheng, editors, *Discrete Mathematical Chemistry*, volume 51 of *DIMACS series in discrete mathematics and theoretical computer science*, pages 377–383. American Mathematical Society, Providence, RI, 2000.