Digital Discovery

Accepted Manuscript

This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

Journal Name

ARTICLE TYPE

Higher order structures in chemistry: hypergraphs reshape the molecule and the reaction[†]

Guillermo Restrepo[*a]

Chemical systems contain higher-order relationships that exceed the binary constraints of traditional graph-based models. Although graph theory has long supported the digital representation of molecules and reactions, many fundamental chemical phenomena—such as multi-centre bonding, aromaticity, cooperative interactions, and the inherently set-theoretical nature of reactions—escape pairwise encodings. This work introduces hypergraphs and an extended "zoo" of higher-order mathematical structures as a unified framework for modelling both molecular structures and reaction networks. Molecular hypergraphs naturally capture multi-atomic interactions, while directed hypergraphs offer a mathematically faithful representation of reactions as transformations between arbitrary sets of substances. More sophisticated variants—including ordered, directed, binary, and directed-ordered hypergraphs—enable the incorporation of additional chemical information, such as atomic ordering, ligand–pocket affinities, and cavity organisation in porous materials at the substance level, as well as toxicity and economic constraints at the reaction level. Recent advances in hypergraph spectral theory, random models, and higher-order network statistics have opened new chemical, mathematical, and computational avenues. These developments coincide with emerging machine-learning evidence showing that hypergraph-based representations of molecules can outperform graph-based and even 3D-coordinate models. By outlining both the capabilities and current limitations of hypergraph approaches, this work argues that higher-order mathematical structures will be central to the next generation of digital discovery, enabling more faithful representations of chemical complexity and deeper integration across chemistry, mathematics, and computer science.

1 Introduction

Chemistry is characterised by its concern with the organisation, behaviour, and transformation of matter at multiple scales. Throughout its history, chemists have relied on high levels of abstraction to interpret, organise, and systematise the practice of their discipline.[1] From molecular models, encoding atomic relationships, to the periodic system, including sets and order relationships,[2] abstract frameworks have long been essential in advancing chemical knowledge.

Building on this historical interplay between chemistry and mathematical abstraction, the present Perspective introduces hypergraphs as a mathematical structure of particular relevance for addressing chemical questions at both the molecular and reaction levels. In doing so, it aims to broaden the ongoing dialogue between chemistry and mathematics and to point towards new directions for chemical theory, computation, and practice.

2 Higher order structures

Gerhardt, the visionary 19th-century chemist,[3–5] produced scientific results that addressed two foundational pillars of chemistry. Through his type representation of substances, he helped lay the groundwork for what later became molecular structural theory,[4] a conceptual framework that remains central to the language of chemistry and to the way we encode chemical knowledge for computational analysis.[3] Gerhardt also contributed to the modern representation of chemical reactions in the now-standard form A + B → C + D.[i] Viewed from a contemporary mathematical perspective, these contributions highlight the fundamental role of relationships in chemistry, from the atomic to the molecular level, that is form substances to reactions.

The second half of the 19th century further strengthened the connection between chemistry and the mathematics of relations. Sylvester, a leading 19th-century mathematician, recognised a connection between the *molecular structures* of his contemporary

---

[a] *Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, 04103 Leipzig, Germany. Fax: +49 (0) 341 9959 658; Tel: +49 (0) 341 9959 601; E-mail: restrepo@mis.mpg.de*

[†] To the memory of Rainer Brüggemann, pioneer on the theory and application of partially ordered sets to chemistry and environmental sciences.

[i] A notation whose roots, as some historians have argued, can be traced back to Lavoisier at the end of the 18th century.[6]

chemists and what mathematicians would later call *graphs*.[7,8] In this connection, atoms correspond to vertices and chemical bonds to edges (pairs of vertices), a correspondence Sylvester himself termed "chemicographs."[8] This insight established a lasting bridge between chemical structural theory and mathematical graph theory,[9,10] which has generated a large amount of research and has led to applications ranging from structural prediction to QSAR (Quantitative Structure-Activity Relationships) modelling.[1,11–13]

Similar ideas emerged when chemists began to use graphs to model the reactivity of substances, ultimately leading in the 20th century to the study of reaction networks, in which substances are represented as vertices and their reactivity relations as edges.[14–17] This mathematical framework is at the basis of the multiple studies on metabolic networks and on the gigantic network spanning the chemical space.[14,15,18]

Despite these successes, graph-based models face intrinsic limitations. As discussed in the next section, not all molecules can be reduced to binary atomic relations. Aromatic compounds, multicentre bonds, organometallic species, and numerous other systems require the interaction of sets of atoms rather than simple pairs. Whenever two sets of atoms interact—and especially when at least one of those sets contains multiple atoms—graphs, in the strict sense, are insufficient.

Chemical reactions exhibit a parallel limitation: they are fundamentally set-theoretical entities. A reaction relates a set of educts to a set of products, and each of these sets frequently contains more than one substance. Representing such transformations through binary edges inevitably leads to the loss or distortion of essential chemical information.

Before turning to the next section, it is therefore important to emphasise that chemistry is fundamentally shaped by higher-order relationships. The discipline extends far beyond binary interactions: both molecular structure and chemical reactivity routinely depend on the collective behaviour of atom sets and substance sets. Higher-order relations are not rare exceptions but a natural feature of both molecular structure and chemical reactivity, and recognising them is essential for developing mathematical frameworks that truly reflect chemical reality.

2.1 Graphs, hypergraphs and the molecular structure

This section outlines the achievements and constraints of chemical graph theory before turning to hypergraphs as a natural generalisation capable of encoding higher-order atomic relations.

2.1.1 Molecular graphs

A molecule is a collection of atoms standing in specific relationships to one another, and graphs offer a straightforward model for such relational structure. Formally, a *graph* $G = (V,E)$ consists of a set of *vertices* V, representing atoms, and a set of *edges* E, representing chemical bonds. For methane, for example, $V = \{C, H^1, H^2, H^3, H^4\}$ and $E = \{\{C, H^1\}, \{C, H^2\}, \{C, H^3\}, \{C, H^4\}\}$ (Figure 1a, where, for simplicity, H atoms are not labelled). Because edges are unordered pairs, the bond $\{C, H^1\}$ is equivalent to $\{H^1, C\}$.
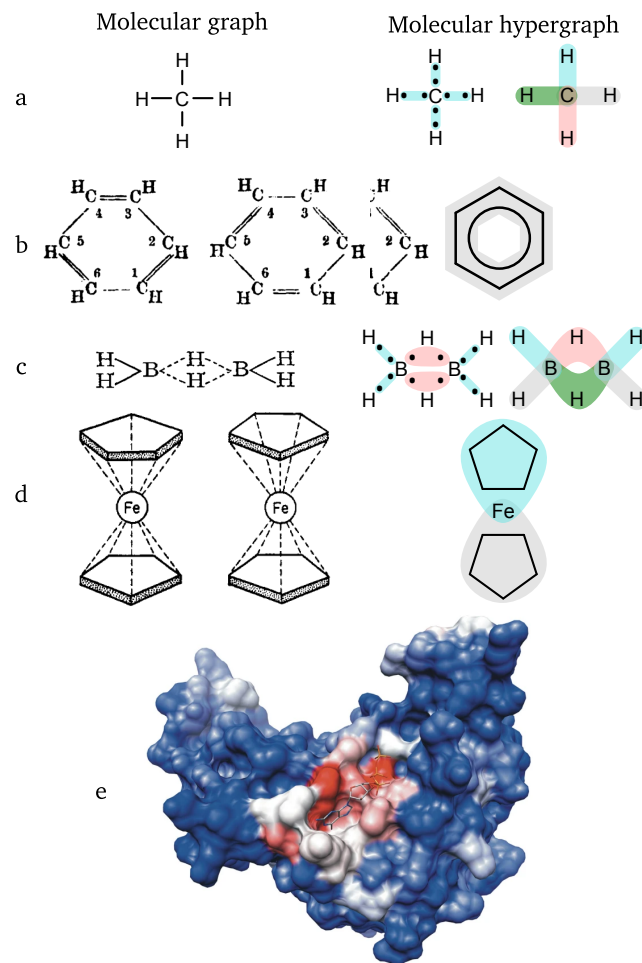

Fig. 1 Graph and hypergraph molecular models. a) Methane, b) benzene, c) diborane, d) ferrocene, e) non-covalent interactions. In a) and c) hyperedges correspond to chemical bonds (shown as coloured regions), with shared electrons (dots). In b) equivalent aromatic carbons are represented by a grey hyperedge. In d) each cyclopentadienyl ring forms a hyperedge describing its collective interaction with the Fe centre. In e) the ligand–protein pocket interaction is represented by hyperedges gathering the highly interacting protein regions (red) together with the corresponding ligand fragments (figure reproduced from [19], published under a CC BY 2.0 license. Reproduced with permission under the Creative Commons Attribution 2.0 International License). Structures in b) (left) as depicted by Kekulé (reproduced from [20] with permission from John Wiley and Sons, copyright 2026). Structure in c) (left) as published by Longuet-Higgins and Bell and reproduced from [21] with permission from The Royal Society of Chemistry, copyright 2026. Structures in d) (left) as depicted by Wilkinson, Rosenblum, Whiting, and Woodward (reproduced from [22] with permission from the American Chemical Society, copyright 2026).

A key strength of molecular graph theory, key for computational encoding and applications, is that graphs can be algebraised. Once encoded as matrices—typically adjacency or incidence matrices—graphs become amenable to linear algebra.[ii]

[ii] The adjacency matrix is an $m \times m$ array indicating whether atom i of the m atoms in the molecule is directly bonded to atom j. A value of 1 in the (i,j) entry denotes the presence of a bond, and 0 its absence. Alternative schemes also exist in which bond orders, rather than simple presence/absence, are recorded. The incidence matrix, by contrast, is an $m \times n$ array encoding the relationship between atoms and bonds: an

This enables the derivation of eigenvalues, eigenvectors, spectra, path counts, cycle detection, and many other mathematically defined quantities with chemical interpretation.[23–25] From these representations emerged the vast field of molecular descriptors, now numbering in the thousands,[26] forming the backbone of QSAR modelling.[1,11–13]

Despite these successes, a fundamental limitation remains: not all molecules can be adequately represented using only binary atomic relations.[27,28] The inadequacy was already evident in the 19th century for benzene and other aromatic systems,[20,29] and re-emerged in the mid-20th century for diborane and ferrocene.[21,22] In modern contexts such as receptor–ligand complexes, binary-only representations are even more restrictive.[30]

In all these cases—aromaticity, multi-centre bonding, organometallic bonding, and protein–ligand recognition—the essential feature is their set-theoretical nature: benzene involves six atoms interacting collectively; the bridging bonds in diborane involve three atoms; each cyclopentadienyl ring in ferrocene interacts as a five-atom unit with Fe; protein pockets interact with a set of ligand atoms (Figure 1). These interactions cannot be reduced without loss to pairs of atoms.

The recognition that molecular structure routinely depends on higher-order atomic relationships lies at the heart of the hypergraph approach. Hypergraphs emerged in the chemical literature in the 1990s,[11,28,31–33] and have recently been rediscovered—both as a faithful mathematical model of multi-atom interactions and as a powerful computational framework.[11,28] One goal of this perspective is to encourage chemists, mathematicians, and computer scientists to revisit hypergraphs as a natural representation of molecular structure.

2.1.2 Molecular hypergraphs

Mathematicians have long studied structures capable of encoding relationships that go beyond pairwise interactions.[34,35] One such structure is the hypergraph,[iii] which extends a graph by allowing its edges—now hyperedges—to contain an arbitrary number of vertices. The number of vertices in a hyperedge—its *size*—may range from one to all m vertices of the hypergraph. This flexibility contrasts with the strict size-2 edges of graphs.

Formally, a *hypergraph* is defined as $H = (V, E)$, where V is the set of *vertices* and E the set of *hyperedges*. A hyperedge may be any subset of V. For instance, a hypergraph model of diborane B_2H_6 can be written as $V = \{\{B^1, B^2, H^1, H^2, H^3, H^4, H^5, H^6\}$ and $E = \{\{B^1,H^1\}, \{B^1,H^2\}, \{B^1,H^3,B^2\}, \{B^1,H^4,B^2\}, \{B^2,H^5\}, \{B^2,H^6\}\}$, where the two three-atom hyperedges B^1,H^3,B^2 and B^1,H^4,B^2 encode the two bridging 3-centre–2-electron bonds (Figure 1c, where atoms are not labelled for simplicity). Note how these hyperedges cannot be represented in a graph setting, since they contain more than two vertices.

Parallel representations for methane, benzene, and ferrocene are shown in Figure 1.

As with molecular graphs, the usefulness of hypergraphs becomes most apparent when they are algebraically encoded so that computations can be performed. In analogy with graphs, hypergraphs can be represented by adjacency matrices, where each entry records whether the i-th and j-th atoms participate together in at least one hyperedge. Here, "participation in a relation" generalises the notion of a chemical bond: for diborane, for example, the adjacency matrix records the three-atom relations of the bridging {B, H, B} bonds, the cyclic six-atom relation in benzene, and the five-atom cyclopentadienyl rings of ferrocene. Hypergraphs also admit incidence matrices, recording vertex membership in each hyperedge.

Based on these matrix representations, spectral hypergraph theory has begun to emerge, encompassing Laplacians, eigenvalues, eigenvectors, and associated invariants.[40–46] Remarkably, the connection to chemistry was already recognised in the 1990s, when the first hypergraph-based molecular descriptors were proposed.[31–33] These developments parallel the trajectory of molecular graph theory, where algebraic descriptors became central to QSAR research.

Recent machine learning work further strengthens the case for hypergraph formulations. Hypergraph neural networks have shown clear advantages over graph-based models—and even over some 3D-coordinate–based representations—in predicting molecular properties and reactivity.[28] These methods leverage the ability of hypergraphs to encode multi-atom interactions natively rather than inferring them indirectly from pairwise data.

*

The trajectory that once took graph theory from a mathematical curiosity to a central tool of theoretical and computational chemistry now appears to be unfolding for molecular hypergraphs. Their capacity to encode higher-order atomic interactions makes them a natural extension of chemical modelling. With the rapid development of hypergraph mathematics and computation, this framework offers a fertile arena for future collaboration between chemists, mathematicians, and computer scientists.

Yet chemistry encompasses both substances and their transformations. Just as hypergraphs enrich the modelling of molecular structure, they also provide expressive representations of chemical reactions and reaction networks. This is the focus of the next section.

2.2 Graphs, hypergraphs and chemical reactions

The algebraic representation of chemical reactions, developed since Gerhardt's times,[3] is now commonly interpreted through graph-theoretical lenses. Graphs have enabled large-scale analyses of reactivity patterns, retrosynthetic pathways, and chemical networks.[14–17] Yet, as with molecular structure, the binary constraints of graphs impose limitations on the faithful encoding of chemical transformations. This section outlines the possibilities and shortcomings of graph-based reaction models and shows how higher-order structures—specifically directed hypergraphs—provide the appropriate mathematical generalisation.

---

entry of 1 in row i, column j indicates that atom i is incident to (that is, participates in) bond j; otherwise the entry is 0. Further details on these matrix representations of molecular graphs may be found in [23].

iii Other higher-order frameworks include simplicial complexes, which have also found chemical applications through topological data analysis.[11,36–38] Further structures include tensors and higher-order Markov chains.[39]

Digital Discovery Accepted Manuscript

2.2.1 Reaction graphs

Over the years, several mathematical frameworks have been proposed to model chemical reactions.[14–18,27] Some describe the dynamics of concentration changes during reaction progress, whereas others focus on the relational structure linking substances and their roles as educts and products. The latter has been the natural domain for graph-theoretical modelling, which constitutes the underlying structure for studying reaction kinetics in large chemical systems.[18]

One of the simplest and most widely used approaches is the *educt–product* model.[27iv] In this representation, a transformation in which substance B is formed from A is encoded as A → B. Formally, such a relation is represented as an *arc*, or *directed edge*, (A, B). This construction has been adopted extensively in studies of metabolic, synthetic, and general chemical reaction networks.[14–18]

However, the model suffers from a fundamental limitation inherent to its binary nature.[27,47–49] Consider the reaction A + B → C + D. The educt–product model represents this transformation as the four arcs A → C, A → D, B → C, B → D, yielding the graph $G = (V, E)$ with $V = \{A, B, C, D\}$ and $E = \{(A,C), (A,D), (B,C), (B,D)\}$ (see reaction r_1 in Figure 2).

$$r_1: A + B \longrightarrow C + D$$
$$r_2: E + F \longrightarrow A$$
$$r_3: C + H \longrightarrow E + G$$
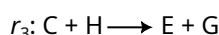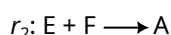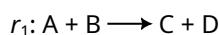
Reaction graph
(directed graph)
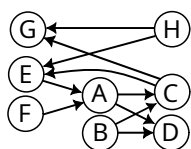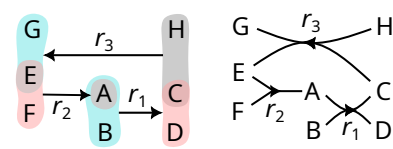
Reaction hypergraph
(directed hypergraph)

Fig. 2 Chemical reaction models. Directed graph and directed hypergraph representations of the three reactions shown at the top. In the graph-based model, educts and products are connected by directed edges (arrows), each encoding a binary relation. In contrast, the hypergraph model relates sets of educts to sets of products via directed hyperedges, thereby capturing the set-theoretical structure of reactions. The arrow-based directed hypergraph notation follows Figure 1 in [50].

As discussed in [27], recovering reactions from this graph leads to multiple spurious interpretations, all of which are consistent with the graph but inconsistent with chemistry. From G, one may infer not only the true reaction A + B → C + D, but also unimolecular reactions A → C, A → D, B → C, and B → D, as well as the artificial bimolecular reactions A + B → C and A + B → D. These arise because the model decomposes each true reaction into pairwise relations between individual substances, ignoring the essential fact that educts act jointly.

The consequences become even more problematic in network-level analyses. In the network shown in Figure 2, the graph model

incorrectly predicts that substances D and G remain reachable even when A is absent, owing to paths such as B → D, B → C → G. Yet the underlying chemistry—embodied in the reactions at the top of the figure—shows clearly that none of these products can form without A. The root of the problem is conceptual: A chemical reaction does not relate individual substances but rather sets of substances.

Formally, a reaction maps a set of educts to a set of products. The educt–product model only faithfully represents rearrangements, where both sets have cardinality 1. But empirical analysis of the chemical literature reveals that chemists overwhelmingly report reactions involving one to three educts and one or two products.[51] This diversity of set sizes cannot be encoded in the binary framework of graphs.

A mathematically faithful representation of chemical reactions requires a framework that accommodates relations between sets of arbitrary size. This is precisely what directed hypergraphs provide, as discussed in the next subsection.

2.2.2 Reaction hypergraphs

The mathematical framework that naturally captures the set-theoretical structure of chemical reactions is the *directed hypergraph* model—a generalisation of directed graphs such as those used in the educt–product representation. Directed hypergraphs treat reactions as relations between sets of arbitrary size, rather than between individual substances.

Formally, a directed hypergraph $H = (V, E)$ consists of a set of vertices V and a set of hyperedges E. Each hyperedge is an ordered pair of vertex-sets (e_i, e_j), where e_i is the tail (or source) and e_j the head (or target). In chemical terms, e_i represents the set of educts and e_j the set of products. The sets e_i and e_j are often called *hypervertices* or *hypernodes* to emphasise that they generalise the notion of a vertex by potentially containing several vertices.[46]

For example, the reaction A + B → C + D is represented by the directed hypergraph $H = (V, E)$ with $V = \{A, B, C, D\}$ and $E = (\{A, B\}, \{C, D\})$. This contrasts sharply with its graph representation (Section 2.2.1), which requires four arcs: A → C, A → D, B → C, and B → D. Likewise, the three reactions shown in Figure 2 are captured by only three directed hyperedges rather than 10 arcs—illustrating the parsimony and clarity of the hypergraph representation.

These examples show that directed hypergraphs encode the set-theoretical essence of chemical reactions: both the educt set and the product set may contain any number of substances. Importantly, the chemical consequences of the absence of a substance become immediately transparent. As discussed in Section 2.2.1, removing A from the system makes the reaction r_1 impossible, which in turn prevents the formation of C and thus blocks reaction r_3 and the production of G. Such dependencies are obscured in the educt–product model.

As with molecular and reaction graphs, directed hypergraphs admit algebraic representations. Adjacency matrices of dimension $m \times m$ (with m the number of hypervertices) encode whether a hypervertex e_i leads to another hypervertex e_j. Incidence matrices of dimension $m \times n$ (with n the number of reactions) indicate

---

iv In[47] this model is referred to as the "substrate graph."

whether a hypervertex acts as educt (entry $+1$) or product (entry -1) in a given reaction; a zero indicates non-participation. These matricial forms enable the study of the structural and dynamical properties of reaction networks.

Recent developments have extended classical network statistics to directed hypergraphs, including clustering coefficients, spectral measures, curvature, shortest paths, communicability, and random models.[46,49,52–62] Several have already been applied to large biochemical networks. For instance, clustering coefficients reveal that metabolic networks are far less clustered than human-made systems such as email networks.[v] Spectral centralities have likewise been used to study biochemical networks, urban transit systems, and propositional-logic databases.[53]

Thermodynamic constraints have recently been incorporated into directed hypergraph reaction networks to identify pathways composed exclusively of energetically favourable reactions.[63] Despite these advances, further work is needed to unify approaches, identify redundancies, and determine whether chemical networks possess structural signatures distinguishing them from other directed hypergraphs.

A further active avenue of research concerns random models for reaction hypergraphs, analogous to random graph models in classical network theory.[64,65] An Erdős–Rényi–type model has been proposed for chemical hypergraphs,[46] enabling comparisons between empirical networks and suitable null models.[vi] Yet many well-known random-graph models—including Watts–Strogatz small-world[66] and Barabási–Albert preferential-attachment models[67]—remain to be generalised to the hypergraph setting. A key variable unique to hypergraphs, and absent from graph-based studies, is *hypervertex size*. This additional degree of freedom is essential and must be incorporated into any statistical or generative model of reaction networks.

<center>*</center>

In summary, directed hypergraphs provide a mathematically expressive and chemically faithful framework for encoding reactions and reaction networks. Their recent mathematical development opens many avenues for future research at the interplay of chemistry, mathematics, and computer sciences, offering fresh perspectives on the structure and organisation of large chemical networks—including those spanning the full chemical space. The following section extends the discussion to further hypergraph structures of chemical relevance.

2.3 A zoo of hypergraph structures of chemical application

In this section, I introduce a family of mathematical structures obtained by generalising different aspects of graphs, and I outline their potential—and in some cases already demonstrated—applications in chemistry.

2.3.1 A zoo of higher-order structures

Graphs encode binary relationships between vertices, that is, relationships between single-vertex sets. In graph theory, edges correspond to unordered pairs of vertices (Figure 3). Thus, the membership of a vertex in an edge signifies its relation to exactly one other vertex. If this notion of membership is relaxed so that a vertex may be related to an arbitrary number of other vertices, the resulting structure is a *hypergraph*: a collection of subsets of vertices. If, in addition, binary relationships are allowed between such vertex-sets of arbitrary size, one obtains a *binary hypergraph* (Figure 3).[vii] By restricting each hypervertex in a binary hypergraph to contain a single vertex, one recovers an ordinary graph. All structures in Figure 3 take the abstract form $G = (V, E)$, with V denoting the underlying vertex set while E varies according to the specific structure.

Beyond graphs, hypergraphs, and binary hypergraphs, richer structures emerge once order relations are introduced. A familiar case is the *directed graph*, obtained by endowing each two-element edge with a direction, thereby producing arcs (Figure 3). If sets of arbitrary size are allowed while retaining an internal order structure, *ordered hypergraphs* are obtained.[viii] Ordered hypergraphs thus constitute collections of partially ordered sets (*posets*)[71] over V. If, in addition, order relations are permitted between pairs of these posets, the result is a *directed ordered hypergraph* (Figure 3). Removing the internal order within hypervertices yields a *directed hypergraph*,[57,72] while removing direction but retaining internal ordering leads to an *ordered binary hypergraph*. Adding direction to the latter reconstructs the directed ordered hypergraph. These interrelations are examples of morphisms between higher-order structures, some of which are illustrated in Figure 3.

Although often not explicitly recognised as such, many chemical systems already exhibit the defining features of these higher-order structures. In the remainder of this section, I discuss several examples and highlight how exploring the mathematics of these structures may deepen our understanding of chemicals and reactivity.

2.3.2 A zoo of higher-order structures in chemical action

Graphs and directed graphs have long been central to chemical modelling, as discussed in earlier sections. Graphs constitute the traditional mathematical representation of molecular structure (Figure 1a), while directed graphs refine this description by encoding inductive effects, bond polarisation and electron flow (Figure 4a), as well as non-covalent interactions such as hydrogen bonding (Figure 4b).[73] Graphs and directed graphs have also been widely used to represent reaction networks. Directed

---

[vii] As shown in Figure 3, binary hypergraphs can be obtained from directed hypergraphs by disregarding the order between hypervertices. Binary hypergraphs preserve the binary relations between hypervertices and may also be viewed as hypergraphs whose hyperedges are partitioned into two subsets of vertices. In this sense, they could be treated as bipartite hypergraphs.

[viii] Ordered hypergraphs are often defined as hypergraphs endowed with a total order on their vertices.[68,69] Here, following the spirit of[70], I consider hypervertices as partially ordered sets.

Graphs $G = (V, E)$
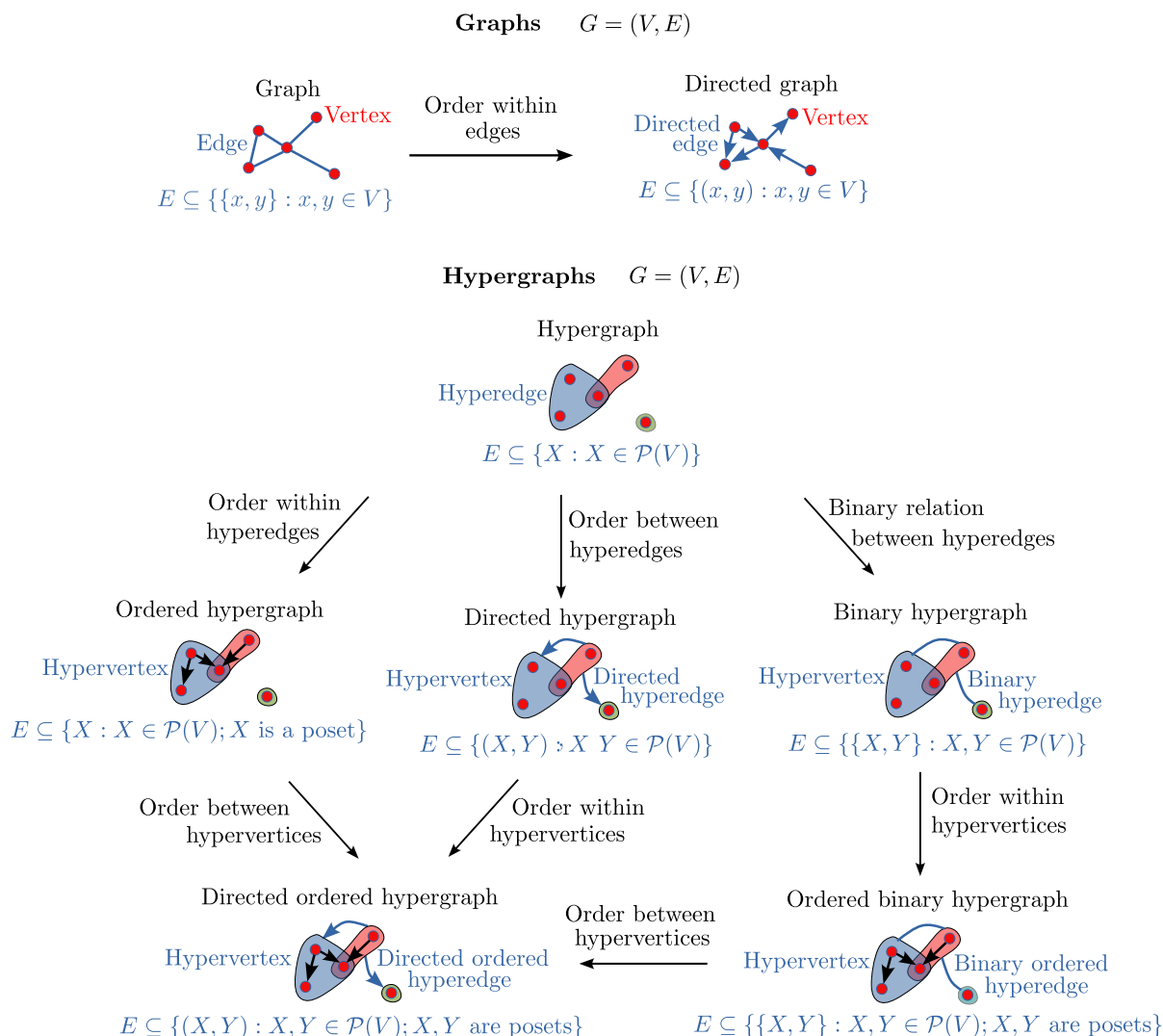

Hypergraphs $G = (V, E)$

Fig. 3 Graphs, hypergraphs, and selected morphisms. All structures are defined as $G = (V, E)$, where V is the set of vertices (red dots) and E collects the specific subsets of vertices associated with each type of structure. In every case, $\mathscr{P}(V)$ denotes the power set of V. Blue lines and arrows represent (hyper)edges, that is the elements of E. Black arrows indicate partial-order relations within the host sets, which thereby form partially ordered sets (posets). Morphisms between structures are shown as arrows connecting them, with the main transformation characterising each morphism indicated alongside.

graphs distinguish educts from products by the orientation of arcs, whereas undirected graphs remain agnostic to this distinction—useful when only the connectivity pattern of a reaction network is of interest rather than the roles of substances.

Hypergraphs offer a natural way to reduce and navigate chemical complexity. The chemical space contains millions of known substances;[51] classification is therefore essential.[50,79] Classes of compounds or classes of reactions correspond to hyperedges in a hypergraph of the chemical space. Alkali metals, halogens, endocrine disruptors, and metal–organic frameworks (MOFs) are examples of hyperedges. Likewise, reaction classes—such as amide formation, Diels–Alder cycloadditions, Buchwald–Hartwig couplings—form another family of hyperedges. These sets often intersect, as molecules or reactions commonly belong to multiple classes, yielding a hypergraph with non-disjoint hyperedges.

At the molecular scale, hypergraphs arise naturally when atomic equivalence classes or multi-centre bonding motifs are used to define hyperedges. Any molecule can thus be modelled as a hypergraph whose hyperedges encode general m-centre–n-electron bonding patterns (Figure 1).

Ordered hypergraphs are equally widespread. Ordering chemical elements by atomic radius or electronegativity yields ordered hypervertices within the hypergraph of the periodic table. In this setting, hypervertices correspond to families of chemical elements. When several parameters are used simultaneously to characterise chemical elements, posets arise, reflecting cases where no total order exists.[2ix] Toxicity rankings, ligand–pocket affinities, electrochemical series, or spectrochemical series all constitute ordered hypergraphs. At the molecular level, bonds may be

ix In formal terms a poset is defined as a set V endowed with an order relation $\preceq$. This later is a binary relation satisfying that for all $x, y, z \in V$: 1) $x \Rightarrow x \preceq x$ (reflexivity), 2) $x \preceq y$ and $y \preceq x \Rightarrow x = y$ (antisymmetry), and 3) $x \preceq y$ and $y \preceq z \Rightarrow x \preceq z$ (transitivity).[71]
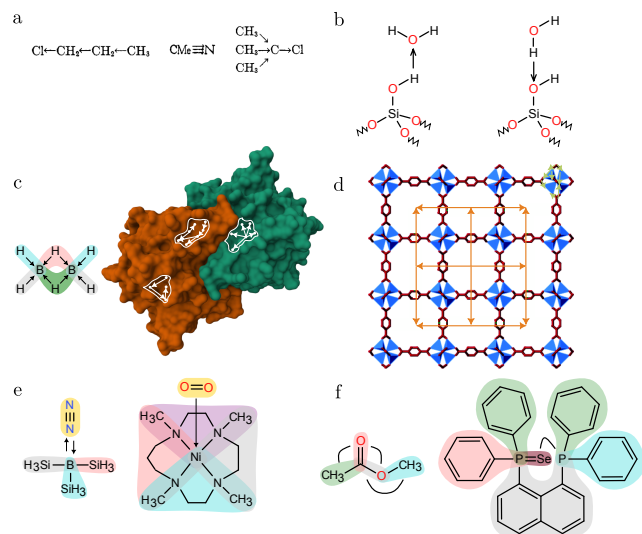
Fig. 4 Graphs and hypergraphs as molecular structure models. Directed graphs representing a) inductive effects as depicted by Ingold (reproduced from [74] with permission from the American Chemical Society, copyright 2026) and b) hydrogen bonds (adapted from Figure 7 in [73] with permission from the Royal Society of Chemistry, copyright 2026). c) Ordered hypergraphs encoding electronegativity differences within bonds (left), and encoding the poset structure (white arrows) of specific regions (white dots) inside protein pockets (white curves) (right). d) Directed ordered hypergraph on a 2D layer of MOF-5 (adapted from Figure 1 in [75] with permission from the the Royal Society of Chemistry, copyright 2026). Cavities (hypervertices) are ordered according to host–guest affinity (orange arrows), and internal hypervertex order is shown at top right, which illustratively indicates the exposition of some atoms to the cavity. e) Directed hypergraphs encoding σ-donation and π-backdonation between $B(SiH_3)_3$ and N_2 [76] (left), and the side-on interaction between O_2 and Ni in $[Ni^{III}(13\text{-}TMC)(O_2)]^+$ (right), with 13-TMC as the N-tetramethylated cyclam chelate with 13 atoms of ring size. [77] f) Binary hypergraphs depicting bond–bond interactions in an ester (left) and Se–P spin–spin coupling in $1\text{-}Ph_2P(C_{10}H_6)\text{-}8\text{-}P(:Se)Ph_2$ (right). [78]

ordered by electronegativity difference or by polarisability (Figure 4c). Protein pockets are also amenable of being modelled using ordered hypergraphs, where products contain regions ordered by their interaction strength with specific ligands (Figure 4c). MOFs accept a similar treatment, where cavities may be ordered by their host–guest affinity. [80,81]

The periodic system, besides being the icon of chemistry, is a chemical object with a rich mathematical structure. [2,82] Here, the corresponding hypergraph structure is the directed ordered hypergraph (Figure 3), where hypervertices are element families endowed with internal order (for instance, by atomic size or electronegativity), while hypervertices are themselves ordered, as clearly recognised in group trends such as alkali metals being more electropositive than halogens. A further case of a chemical directed ordered hypergraph is the resulting from ordering chemicals by their degree of substitution. [83,84] In this case, hypervertices are chemicals with the same degree of substitution.[x]

Directed ordered hypergraphs also naturally arise in catalyst selection. Suppose catalysts are classified into Pd-based and Ni-based families (hypervertices), with overlapping members. Additional parameters—cost, toxicity, availability—induce internal poset structures. Directed ordered relations between hypervertices enable the selection of the most promising catalyst class, after which internal ordering identifies the optimal candidate. Therefore, exploring the mathematical properties of this model may improve the AI-driven studies on synthesis planning under realistic chemical and external constraints.

The same framework supports retrosynthetic analysis and decision-making in self-driving laboratories. [85] Substances form posets based on several criteria (toxicity, solubility, cost), reactions form hypervertices, and directed order between hypervertices identifies preferable synthetic routes.

Directed ordered hypergraphs also model the ordering of cavities within MOFs (Figure 4d). Cavities (hypervertices) are ordered by their adsorption capacity or accessibility, while their internal order reflects steric and electronic factors arising from secondary building units.

Back to Figure 4, a further hypergraph structure is the directed hypergraph, which, as shown earlier, provides a rigorous framework for modelling chemical reactions. [46,51] They also offer molecular-level insight, for example by encoding directed interactions between atom sets, as in σ-donation and π-backdonation in metal complexes or the side-on binding of O_2 to transition metals (Figure 4e). [76,77,86]

Binary hypergraphs, in turn, capture interactions between sets of atoms and are particularly relevant for bond–bond couplings. Therefore, these hypergraph structures constitute suitable molecular models of application in different spectroscopies. For instance, vibrational spectroscopies such as IR and Raman, which reveal mode couplings—for example, between C=O and C–O stretches in esters (Figure 4f)—could find a natural framework for interpretation with binary hypergraphs, where interacting bonds become the centre of attention. Likewise, the long-range spin–spin couplings (Figure 4f), detected with NMR spectroscopy find in binary hypergraphs a suitable model.

Chemical reaction networks can also be modelled as binary hypergraphs when the goal is to study global connectivity rather than the educt–product distinction. This binary hypergraph model was used recently for developing a random model for the chemical space. [46]

Finally, when atoms or groups of atoms are characterised by multiple parameters—connectivity, conformational state, vibrational signatures, mechanistic role, stereochemistry, electronic structure—they form posets. Interactions between such posets are central in diverse chemical phenomena: coupling between conformational substates during protein folding; [87] vibrational poset interactions through anharmonic couplings; [88] mechanistic posets constraining multistep pathways; [89] stereochemical posets governing diastereomer stability; [90] and electronic-structure posets whose interactions give rise to conjugation or

---

[x] Interestingly, since the order relation between chemicals is given by the embedding of subgraphs of one chemical structure within another, vertices in hypervertices are not comparable, which also equates this chemical system with a directed hyper-

graph. See [83] for further details.

aromaticity.[91] In all these cases, chemical behaviour emerges from the interplay of multiple interacting posets, each encoding a distinct facet of molecular organisation. These cases are suitably modelled by ordered binary hypergraphs (Figure 4).

Conclusions

The modelling of chemical systems at the molecular and reaction levels has traditionally relied on graphs. While these structures have proved extremely fruitful, they also exhibit clear limitations. Graphs are intrinsically binary, and therefore struggle to capture the multi-centric, higher-order, and context-dependent interactions characteristic of both molecular structure and chemical reactions. Hypergraphs, by contrast, naturally encode relations among sets of arbitrary size and therefore provide a far richer mathematical language for representing chemical complexity.

In addition to presenting the uses of some hypergraph structures in chemistry, namely molecular hypergraphs and directed hypergraphs as a model for chemical reactions, a zoo of other hypergraph structures is presented, from ordered hypergraphs to directed ordered hypergraphs and ordered binary hypergraphs, which nuance the description of molecules and chemical reactions.

Despite their advantages, hypergraph-based models must be used with appropriate caution. Like graphs, these higher-order structures capture selected aspects of chemical systems—specifically their relational or topological organisation—but they do not, by themselves, encode all chemically relevant information. Hypergraphs significantly enrich the representation of multi-atomic and multi-molecular relationships, yet whenever three-dimensional geometric details are essential—such as quantifying cavity sizes in MOFs, characterising protein pockets, or distinguishing molecular conformations—additional structural information must be incorporated. This need mirrors the well-known limitations of graph-based molecular representations: embedding topological structures into metric space requires complementing the (hyper)graph with weighted vertices and edges and, crucially, with explicit coordinate systems. In such settings, hypergraphs serve as a powerful relational scaffold, but geometry must be supplied through appropriate metric or spatial augmentations.

A further element of caution when using hypergraphs in chemistry is related to the relative novelty of the hypergraph literature. Hypergraph theory is still an emerging and rapidly developing field, and its nomenclature has not yet reached the level of standardisation enjoyed by graph theory. As a consequence, the same mathematical object may carry different names in different subfields, while distinct objects may be referred to by similar terminology. For this reason, reading the hypergraph literature requires close attention to definitions rather than reliance on terminology alone. This conceptual heterogeneity is a challenge, but also a reflection of the vibrant and expanding nature of the field.

From a computational perspective, the situation is equally nuanced. Several hypergraph statistics can be mapped into statistics on bipartite graphs, which sometimes permits the transfer of efficient graph-theoretical algorithms. Yet this reduction is not universally applicable.[47] In fact, many core tasks in hypergraph analysis remain computationally demanding. For example, computing shortest paths in weighted hypergraphs is NP-hard.[47] Further examples of hypergraph computational complexity are discussed in[92–94]. These challenges have direct implications for chemistry, where algorithmic efficiency is crucial. A case in point is the determination of whether a molecular fragment occurs within a larger structure—a task that reduces to subgraph matching in graph theory but becomes considerably more complex in hypergraphs.[95,96] Understanding and addressing these complexities is essential for the practical use of hypergraphs in chemistry.

The mathematical and computational study of hypergraphs is, as shown throughout this Perspective, an active and rapidly evolving domain. When combined with the richer and more sophisticated hypergraph structures introduced here—structures that extend well beyond classical hypergraphs into what is dubbed as a "zoo" of higher-order relational frameworks—the challenges are amplified. There remains a substantial amount of theoretical and algorithmic work to be done. Addressing these issues is not only a promising direction for mathematics and computer sciences, but also one from which chemistry stands to benefit enormously.

Another frontier concerns machine learning. Recent studies have already shown that hypergraph representations of molecular systems outperform both traditional graph-based models and models incorporating full three-dimensional information.[28] These advances derive from the extension of graph neural networks into hypergraph neural networks,[28] which can exploit higher-order structural information unavailable to pairwise models. It is natural to ask what would happen if the novel hypergraph structures introduced in this paper—ordered hypergraphs, directed hypergraphs, and other enriched forms—were incorporated into new machine learning architectures. Such developments could radically expand the representational and predictive power of models for chemical discovery.

In summary, hypergraphs and their mathematical extensions offer a powerful framework for modelling the richness and complexity of chemical systems. Their integration with modern computational approaches, including algorithmic advances and machine learning, opens a path toward a new generation of tools for digital discovery. By embracing these higher-order structures, chemistry gains access to a deeper and more expressive mathematical language—one capable of capturing complexity that lies beyond the reach of traditional, pairwise models.

Author contributions

G.R. was responsible for conceptualisation, formal analysis, investigation, methodology development, visualisation, and manuscript writing.

Conflicts of interest

There are no conflicts to declare.

Data availability

There is no original data associated to this document.

Acknowledgements

Notes and references

1 G. Restrepo, *Essays in the Philosophy of Chemistry*, Oxford University Press, 2016.

2 W. Leal and G. Restrepo, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2019, **475**, 20180581.

3 C. Gerhardt, *Traité de chimie organique*, F. Didot frères, 1860.

4 H. Chang, *Is Water H_2O?: Evidence, Realism and Pluralism*, Springer Netherlands, 2012.

5 A. J. Rocke, *Chemical Atomism in the Nineteenth Century*, Ohio State University Press, 1984.

6 C. Gillispie, *The Edge of Objectivity: An Essay in the History of Scientific Ideas*, Princeton University Press, 1973.

7 J. J. Sylvester, *American Journal of Mathematics*, 1878, **1**, 64–104.

8 J. J. Sylvester, *Nature*, 1878, **17**, 284–284.

9 G. Restrepo, *Foundations of Chemistry*, 2013, **15**, 185–197.

10 D. J. Klein, *Hyle*, 2013, **19**, 35–85.

11 L. S. G. Leite, S. Banerjee, Y. Wei, J. Elowitt and A. E. Clark, *WIREs Computational Molecular Science*, 2024, **14**, e1729.

12 P.-Y. Cao, Y. He, M.-Y. Cui, X.-M. Zhang, Q. Zhang and H.-Y. Zhang, *Journal of Cheminformatics*, 2024, **16**, 133.

13 H. Qin, A. F. Hashem, M. F. Hanif and O. A. Fiidow, *Scientific Reports*, 2025, **15**, 28025.

14 M. Fialkowski, K. J. M. Bishop, V. A. Chubukov, C. J. Campbell and B. A. Grzybowski, *Angewandte Chemie International Edition*, 2005, **44**, 7263–7269.

15 K. J. M. Bishop, R. Klajn and B. A. Grzybowski, *Angewandte Chemie International Edition*, 2006, **45**, 5348–5354.

16 S. Fujita, *Computer-Oriented Representation of Organic Reactions*, Yoshioka Shoten Publishing Company, 2001.

17 S. Rao, A. van der Schaft and B. Jayawardhana, *Journal of Mathematical Chemistry*, 2013, **51**, 2401–2422.

18 H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabási, *Nature*, 2000, **407**, 651–654.

19 V. Bianchi, P. F. Gherardini, M. Helmer-Citterich and G. Ausiello, *BMC Bioinformatics*, 2012, **13**, S17.

20 A. Kekulé, *Justus Liebigs Annalen der Chemie*, 1872, **162**, 77–124.

21 H. C. Longuet-Higgins and R. P. Bell, *J. Chem. Soc.*, 1943, 250–255.

22 G. Wilkinson, M. Rosenblum, M. C. Whiting and R. B. Woodward, *Journal of the American Chemical Society*, 1952, **74**, 2125–2126.

23 N. Trinajstić, *Chemical Graph Theory*, CRC Press, 2018.

24 N. Biggs, *Algebraic Graph Theory*, Cambridge University Press, 1993.

25 X. Li, Y. Shi and I. Gutman, *Graph Energy*, Springer, 2012.

26 R. Todeschini, V. Consonni, R. Mannhold, H. Kubinyi and G. Folkers, *Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical Listing / Volume II: Appendices, References*, Wiley, 2009.

27 G. Restrepo, *Digital Discovery*, 2022, **1**, 568–585.

28 J. Chen and P. Schwaller, 2024, **160**, 144307.

29 A. J. Rocke, *Angewandte Chemie International Edition*, 2015, **54**, 46–50.

30 K. A. Murgas, E. Saucan and R. Sandhu, 2022, **12**, 20879.

31 E. V. Konstantinova and V. A. Skoroboratov, 1998, **39**, 958–966.

32 X. Liu, L. Wang and X. Li, 2020, **39**, 351–364.

33 E. V. Konstantinova and V. A. Skorobogatov, 1995, **35**, 472–478.

34 P. Corsini and V. Leoreanu, *Applications of Hyperstructure Theory*, Springer US, 2003.

35 C. Berge, *Graphs and Hypergraphs*, North-Holland Publishing Company, 1973.

36 J. Bilsky and A. E. Clark, *The Journal of Chemical Physics*, 2025, **163**, 090901.

37 K. Xia, X. Feng, Y. Tong and G. W. Wei, *Journal of Computational Chemistry*, 2015, **36**, 408–422.

38 K. Xia and G.-W. Wei, *Journal of Computational Chemistry*, 2015, **36**, 1502–1520.

39 A. R. Benson, D. F. Gleich and D. J. Higham, *Higher-order Network Analysis Takes Off, Fueled by Classical Ideas and New Data*, 2021, https://arxiv.org/abs/2103.05031.

40 R. Mulas, C. Kuehn, T. Böhle and J. Jost, *Discrete Applied Mathematics*, 2022, **317**, 26–41.

41 R. Mulas, D. Horak and J. Jost, in *Graphs, Simplicial Complexes and Hypergraphs: Spectral Theory and Topology*, ed. F. Battiston and G. Petri, Springer International Publishing, Cham, 2022, pp. 1–58.

42 J. Jost, R. Mulas and D. Zhang, *Vietnam Journal of Mathematics*, 2022, **50**, 323–358.

43 R. Mulas, C. Kuehn and J. Jost, *Physical Review E*, 2020, **101**, 062313.

44 E. Andreotti and R. Mulas, *Electronic Journal of Graph Theory and Applications*, 2022, **10**, 485–500.

45 R. Mulas and J. Nie, *Acta Mathematica Hungarica*, 2023, **170**, 168–182.

46 A. Garcia-Chung, M. Bermúdez-Montaña, P. F. Stadler, J. Jost and G. Restrepo, *Journal of Mathematical Chemistry*, 2024, **62**, 1357–1383.

47 S. Klamt, U.-U. Haus and F. Theis, *PLOS Computational Biology*, 2009, **5**, 1–6.

48 M. Wen, E. W. C. Spotte-Smith, S. M. Blau, M. J. McDermott, A. S. Krishnapriyan and K. A. Persson, *Nature Computational Science*, 2023, **3**, 12–24.

Digital Discovery Accepted Manuscript

49 C. Flamm, B. M. R. Stadler and P. F. Stadler, in *Generalized Topologies: Hypergraphs, Chemical Reactions, and Biological Evolution*, Bentham-Elsevier, 2015, ch. 2, pp. 300–328.

50 J. Schummer, *Hyle*, 1998, **4**, 129–162.

51 E. J. Llanos, W. Leal, D. H. Luu, J. Jost, P. F. Stadler and G. Restrepo, *Proceedings of the National Academy of Sciences*, 2019, **116**, 12660–12665.

52 G.-G. Ha, I. Neri and A. Annibale, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2024, **34**, 043102.

53 G. Contreras-Aso, R. Criado and M. Romance, *Journal of Complex Networks*, 2024, **12**, cnae037.

54 W. Leal, G. Restrepo, P. F. Stadler and J. Jost, *Advances in Complex Systems*, 2021, **24**, 2150003.

55 M. Eidi and J. Jost, *Scientific Reports*, 2019, **10**, 12466.

56 S. Kim, M. Choe, J. Yoo and K. Shin, *Data Mining and Knowledge Discovery*, 2023, **37**, 2330–2388.

57 G. Ausiello and L. Laura, *Theoretical Computer Science*, 2017, **658**, 293–306.

58 S. S. Shetty and K. A. Bhat, *Arab Journal of Basic and Applied Sciences*, 2025, **32**, 347–360.

59 R. Fagerberg, C. Flamm, R. Kianian, D. Merkle and P. F. Stadler, *Journal of Cheminformatics*, 2018, **10**, 19.

60 P. Traversa, G. Ferraz de Arruda, A. Vazquez and Y. Moreno, *Entropy*, 2023, **25**, year.

61 S. Krieger and J. Kececioglu, *Journal of Computational Biology*, 2023, **30**, 1198–1225.

62 S. G. Marehalli Srinivas, M. Esposito and N. Freitas, *The Journal of Chemical Physics*, 2025, **163**, 194112.

63 A. Pal, R. Fagerberg, J. L. Andersen, C. Flamm, P. Dittrich and D. Merkle, *Journal of Chemical Information and Modeling*, 2025, **65**, 6772–6787.

64 P. Erdős and A. Rényi, *Publicationes Mathematicae Debrecen*, 1959, **6**, 290.

65 M. Newman, *Networks*, OUP Oxford, 2018.

66 D. J. Watts and S. H. Strogatz, *Nature*, 1998, **393**, 440–442.

67 A.-L. Barabási and R. Albert, *Science*, 1999, **286**, 509–512.

68 J. Balogh, B. Bollobás and R. Morris, *Eur. J. Combin.*, 2006, **27**, 1263 – 1281.

69 C. Eslahchi and A. M. Rahimi, *Matematicki Vesnik*, 2007, **59**, 9–13.

70 F. Feng, X. He, Y. Liu, L. Nie and T.-S. Chua, Proceedings of the 2018 World Wide Web conference, Republic and Canton of Geneva, Switzerland, 2018, pp. 1523–1532.

71 W. Trotter, *Combinatorics and Partially Ordered Sets: Dimension Theory*, Johns Hopkins University Press, 2002.

72 M. Thakur and R. Tripathi, *Theoretical Computer Science*, 2009, **410**, 2592–2618.

73 S. Bougueroua, A. A. Kolganov, C. Helain, C. Zens, D. Barth, E. A. Pidko and M.-P. Gaigeot, *Phys. Chem. Chem. Phys.*, 2025, **27**, 1298–1309.

74 C. K. Ingold, *Chemical Reviews*, 1934, **15**, 225–274.

75 P. Cubillas, K. Etherington, M. W. Anderson and M. P. Attfield, *CrystEngComm*, 2014, **16**, 9834–9841.

76 T. Brinck, F. Sagan and M. Mitoraj, *European Journal of Inorganic Chemistry*, 2025, **28**, e202400845.

77 J. Cho, H. Y. Kang, L. V. Liu, R. Sarangi, E. I. Solomon and W. Nam, *Chemical Science*, 2013, **4**, 1502–1508.

78 A. Karaçar, M. Freytag, H. Thönnessen, J. Omelanczuk, P. G. Jones, R. Bartsch and R. Schmutzler, *Zeitschrift für anorganische und allgemeine Chemie*, 2000, **626**, 2361–2372.

79 A. Bernal, E. Llanos, W. Leal and G. Restrepo, *Advances in Mathematical Chemistry and Applications*, Bentham Science Publishers, 2015, pp. 24 – 54.

80 H. Li, M. Eddaoudi, M. O'Keeffe and O. M. Yaghi, *Nature*, 1999, **402**, 276–279.

81 K. Tan, S. Zuluaga, E. Fuentes, E. C. Mattson, J.-F. Veyan, H. Wang, J. Li, T. Thonhauser and Y. J. Chabal, *Nature Communications*, 2016, **7**, 13871.

82 D. Rouvray and R. King, *The Mathematics of the Periodic Table*, Nova Science Publishers, 2006.

83 G. Restrepo and D. J. Klein, *Journal of Mathematical Chemistry*, 2011, **49**, 1311–1321.

84 D. J. Klein and L. Bytautas, *MATCH Communications in Mathematical and in Computer Chemistry*, 2000, **42**, 261–290.

85 G. Restrepo and P. F. Stadler, *ACS Sustainable Chemistry & Engineering*, 2016, **4**, 2191–2199.

86 B. A. Smith and K. D. Vogiatzis, *The Journal of Physical Chemistry A*, 2021, **125**, 7956–7966.

87 H. Ishikawa, K. Kwak, J. K. Chung, S. Kim and M. D. Fayer, *Proceedings of the National Academy of Sciences*, 2008, **105**, 8619–8624.

88 K. K. Liang and A. A. Villaeys, *Physical Review A*, 2010, **82**, 013423.

89 J. L. Andersen, C. Flamm, D. Merkle and P. F. Stadler, *Chemical Transformation Motifs - Modelling Pathways as Integer Hyperflows*, 2017, https://arxiv.org/abs/1712.02594.

90 S. Datta and T. Limpanuparb, *RSC Advances*, 2021, **11**, 20691–20700.

91 R. Grande-Aztazi, E. Matito, J. M. Ugalde and J. M. Mercero, *ChemPhysChem*, 2025, **26**, e202400713.

92 M. Thakur and R. Tripathi, FSTTCS 2004: Foundations of Software Technology and Theoretical Computer Science, Berlin, Heidelberg, 2005, pp. 481–493.

93 M. Bressan, J. Brinkmann, H. Dell, M. Roth and P. Wellnitz, *The Complexity of Counting Small Sub-Hypergraphs*, 2025, https://arxiv.org/abs/2506.14081.

94 G. Csáji, *Theoretical Computer Science*, 2022, **931**, 1–16.

95 Z. Yang, W. Zhang, X. Lin, Y. Zhang and S. Li, *HGMatch: A Match-by-Hyperedge Approach for Subgraph Matching on Hypergraphs*, 2023, https://arxiv.org/abs/2302.06119.

96 Y. Su, Y. Gu, Z. Wang, Y. Zhang, J. Qin and G. Yu, *IEEE Trans. on Knowl. and Data Eng.*, 2023, **35**, 5808–5822.

There is no original data associated to this document.

Digital Discovery Accepted Manuscript