Ptolemy Journal of Chemistry

PSR Press

Article

HyperGraph- and superhypergraph-based molecular models: weighted, rough, neural, and multipolar frameworks

Takaaki Fujita

Independent Researcher, Tokyo, Japan; takaaki.fujita060@gmail.com

Received: 01 January 2026; Accepted: 30 March 2026; Published: 23 April 2026.

Abstract: We present a unified, set–theoretic framework that extends molecular graphs to hypergraphs and superhypergraphs via iterated power sets. We define Molecular Graphs, Molecular HyperGraphs, and Molecular SuperHyperGraphs, and develop four complements over them: *Weighted*, *Rough*, *Neural*, and *Multipolar* frameworks. We prove concise inclusion results—most notably, that the Weighted Molecular SuperHyperGraph strictly contains both Weighted Molecular HyperGraphs and (unweighted) Molecular SuperHyperGraphs—while preserving alternating–path distances under canonical embeddings. Compact examples (e.g., methane, ethanol, acetic acid) illustrate how atoms, bonds, functional groups, and higher–order motifs appear as vertices, hyperedges, and superedges under rank constraints. We also provide implementation–agnostic message–passing rules for variable–arity interactions, enabling property prediction and hierarchical analysis in chemistry and chemical biology. This paper is devoted to theoretical analysis, and it is hoped that quantitative studies by domain experts will be developed in future work.

Keywords: rough molecular graph, molecular graph, weighted molecular graph, superhypergraph, hypergraph, neural networks

1. Introduction

1.1. From graphs to superhypergraphs

Classical graph models represent *binary* structure: an edge records a relation between exactly two vertices [1]. Yet many systems of practical interest—for instance, team-based communication, multi-agent coordination, and shared-resource protocols—are governed by interactions that simultaneously involve three or more entities. Hypergraphs capture such higher-order phenomena by permitting each *hyperedge* to connect an arbitrary nonempty subset of vertices in one incidence relation [2–4].

Higher arity alone is often insufficient: contemporary data and engineered platforms frequently exhibit *hierarchical* organization and *nested* structure (e.g., modules composed of submodules, communities within communities, or layered service stacks). To formalize this type of multi-level incidence, *SuperHyperGraphs* are constructed by iterating the powerset operator on a finite base set, thereby producing successive levels in which "vertices" may themselves be set-valued objects (sets, sets of sets, and so forth) [5,6]. Research on SuperHyperGraphs has remained very active in recent years [7–9]. Table 1 provides a compact comparison of the three modeling layers: graphs, hypergraphs, and n-SuperHyperGraphs.

Table 1. Key distinctions among graphs, hypergraphs, and n-SuperHyperGraphs.

Concept	*Notation*	*Edge family*	*Core extension principle*
Graph [10]	$G = (V, E)$	$E \subseteq \binom{V}{2}$	Edges represent *pairwise* (binary) relations on the vertex set.
Hypergraph [2]	$H = (V, \mathcal{E})$	$\mathcal{E} \subseteq \mathcal{P}(V) \setminus \{\emptyset\}$	A hyperedge may join *any* nonempty group of vertices, modeling variable-arity interactions.
n-SuperHyperGraph [11]	$\mathrm{SHG}^{(n)} = (V, \mathcal{E})$ (on a base set V_0)	$V \subseteq \mathcal{P}^n(V_0) \setminus \{\emptyset\}, \mathcal{E} \subseteq \mathcal{P}(V) \setminus \{\emptyset\}$	Supervertices lie in an n-fold powerset hierarchy and can be *set-valued*; edges are subsets of the supervertex domain, enabling explicitly *nested* and *multi-level* incidence.

Notation. $\mathcal{P}(X) = \{A \mid A \subseteq X\}$, $\binom{V}{2} = \{\{u, v\} \subseteq V \mid u \neq v\}$, and $\mathcal{P}^0(X) = X$, $\mathcal{P}^{k+1}(X) = \mathcal{P}(\mathcal{P}^k(X))$.

These graphs, hypergraphs, and superhypergraphs have been extended using frameworks such as fuzzy [12,13], intuitionistic fuzzy [14,15], vague [16], neutrosophic [9], plithogenic [17], soft [18], rough [19], and weighted models [20], and their applications in decision-making, applied science, and related fields have been actively studied.

1.2. Hypergraph-based molecular modeling and hypergraph neural networks

Graphs, hypergraphs, and SuperHyperGraphs have also been studied in chemistry. By using hypergraphs and SuperHyperGraphs, it becomes possible to represent hierarchical structures that cannot be adequately expressed by ordinary graphs. For example, Molecular Graphs have been extended through hypergraph and SuperHyperGraph frameworks, leading to the study of Molecular HyperGraphs [21–24] and Molecular SuperHyperGraphs.

Moreover, in the fields of artificial intelligence and neural networks, Graph Neural Networks have been extended by means of hypergraphs and SuperHyperGraphs, resulting in the definitions of Hypergraph Neural Networks [25–27] and SuperHyperGraph Neural Networks [28], both of which have been actively studied. In addition, concepts such as Molecular Graph Neural Networks [29–31] and Molecular HyperGraph Neural Networks [32,33] have also attracted growing research interest.

1.3. Our contributions

These observations underscore the value of studying graphs, hypergraphs, and superhypergraphs as a progressive modeling family. Since graph-based formalisms have proved effective across a wide range of applied domains, it is natural to pursue analogous developments in chemistry, where one seeks to represent atoms, bonds, functional groups, and higher-order motifs within a unified incidence framework.

Accordingly, in this paper we introduce *Molecular Graphs*, *Molecular HyperGraphs*, and *Molecular SuperHyperGraphs*, and we equip them with four complementary extensions: *Weighted*, *Rough*, *Neural*, and *Multipolar* frameworks. We establish several succinct inclusion relations; in particular, the *Weighted Molecular SuperHyperGraph* strictly contains both *Weighted Molecular HyperGraphs* and the (unweighted) class of *Molecular SuperHyperGraphs*. Moreover, under canonical embeddings we show that alternating-path distances are preserved. Small illustrative case studies—including methane, ethanol, and acetic acid—make explicit how atoms, bonds, functional groups, and higher-order motifs arise as vertices, hyperedges, and superedges subject to rank constraints. Finally, we provide implementation-agnostic message-passing rules for variable-arity interactions, supporting property prediction and hierarchical analysis in chemistry and chemical biology. This paper is devoted to theoretical analysis, and it is hoped that quantitative studies by domain experts will be developed in future work.

2. Preliminaries

This section gathers the basic notions and notation used throughout the paper.

2.1. Power sets and iterated power sets

For a set S, the *power set* $\mathcal{P}(S)$ is the family of all subsets of S, including both the empty set and S itself. The n-th iterated power set of S is obtained by applying the power–set operation n times in succession, starting from S [34,35].

Definition 1 (Power Set [36])**.** The *power set* of S, written $\mathcal{P}(S)$, is the collection of all subsets of S:

$$\mathcal{P}(S) = \{ X : X \subseteq S \}.$$

Definition 2 (Nonempty Iterated Power Set [37])**.** For any set X, write $\mathcal{P}^*(X) := \mathcal{P}(X) \setminus \{\varnothing\}$. Define the nonempty iterated power sets of S by

$$\mathcal{P}_1^*(S) = \mathcal{P}(S) \setminus \{\varnothing\}, \qquad \mathcal{P}_{k+1}^*(S) = \mathcal{P}^*(\mathcal{P}_k^*(S)) \quad (k \geq 1).$$

Example 1 (Nonempty Iterated Power Set in a Chemical System: Water H_2O)**.** *Base set (atoms).* Let

$$S = \{O,\ H_1,\ H_2\},$$

where O denotes the oxygen atom and H_1, H_2 the two hydrogens in a single water molecule.

First nonempty iterated power set $\mathcal{P}_1^*(S)$: molecular substructures.

$$\mathcal{P}_1^*(S) = \{\{O\},\ \{H_1\},\ \{H_2\},\ \{O,H_1\},\ \{O,H_2\},\ \{H_1,H_2\},\ \{O,H_1,H_2\}\}.$$

Elements of $\mathcal{P}_1^*(S)$ are the *nonempty* atomic fragments: single atoms (e.g. $\{O\}$), atom pairs (e.g. $\{O,H_1\}$, an O–H bond fragment), and the full molecule $\{O,H_1,H_2\}$.

Second nonempty iterated power set $\mathcal{P}_2^*(S) = \mathcal{P}^*(\mathcal{P}_1^*(S))$: families of substructures. Each element of $\mathcal{P}_2^*(S)$ is a *nonempty family* of the nonempty fragments above. For instance:

$$\mathcal{F}_{\text{bonds}} = \{\{O,H_1\},\ \{O,H_2\}\} \in \mathcal{P}_2^*(S),$$

collects the two O–H bond fragments simultaneously (useful as a coarse-grained "both bonds" description). Two further, chemically interpretable families are

$$\mathcal{F}_{\text{whole}} = \{\{O,H_1,H_2\}\}, \qquad \mathcal{F}_{\text{decomp}} = \{\{O\},\ \{H_1,H_2\}\},$$

representing, respectively, the intact molecule (as a singleton family) and a hypothetical decomposition into an oxygen fragment and an H_2 fragment.

Third level $\mathcal{P}_3^*(S)$: families of families (optional). To illustrate nesting, the collection

$$\mathcal{G} = \{\ \mathcal{F}_{\text{bonds}},\ \mathcal{F}_{\text{whole}}\ \} \ \in\ \mathcal{P}_3^*(S),$$

encodes alternative coarse-grainings ("both bonds" vs. "whole molecule") within a single object. Such higher iterates naturally model hierarchical choices among fragmentation schemes, functional views, or analysis tasks.

Use in chemistry/biology. In chemistry, $\mathcal{P}_1^*(S)$ enumerates nonempty fragments (atoms, bonds, whole molecule), while $\mathcal{P}_2^*(S)$ captures *sets of fragments* used together in coarse-grained models, reaction templates, or constraint sets. The same construction applies in biology by letting S be, for example, a set of residues in a protein: $\mathcal{P}_1^*(S)$ lists nonempty residue subsets (motifs), and $\mathcal{P}_2^*(S)$ groups motifs into families (alternative binding-site definitions or multi-motif modules).

2.2. Hypergraphs and superhypergraphs

Hypergraphs extend ordinary graph theory by permitting each *hyperedge* to link an arbitrary nonempty subset of the vertex set, thereby capturing higher–order relations among entities [2,38]. A *SuperHyperGraph* pushes this idea further by iterating the power–set construction so that vertices and edges may themselves be sets drawn from successive powers of a base set, enabling multi–layered and self–referential connectivity patterns [5,11]. Owing to this expressive capacity, SuperHyperGraphs have recently attracted growing interest and applications across several domains [39,40].

Definition 3 (Hypergraph [38,41])**.** Given a finite vertex set V, a *hypergraph* is a pair

$$H = (V,E), \qquad E \subseteq \mathcal{P}(V) \setminus \{\varnothing\},$$

whose elements $e \in E$ are called *hyperedges*. In particular, no bound is placed on the size of a hyperedge other than being nonempty.

Example 2 (Biological Hypergraph: A Metabolic Reaction Fragment)**.** Let the vertex set collect metabolites

$$V = \{\text{Glucose},\ \text{ATP},\ \text{G6P},\ \text{ADP},\ \text{F6P},\ \text{F16BP}\}.$$

Define two hyperedges representing multi–metabolite reactions:

$$e_{\mathrm{HK}} = \{\mathrm{Glucose},\ \mathrm{ATP},\ \mathrm{G6P},\ \mathrm{ADP}\} \quad (\text{hexokinase step: Glucose+ATP} \rightleftharpoons \text{G6P+ADP}),$$
$$e_{\mathrm{PFK}} = \{\mathrm{F6P},\ \mathrm{ATP},\ \mathrm{F16BP},\ \mathrm{ADP}\} \quad (\text{PFK step: F6P+ATP} \rightleftharpoons \text{F1,6BP+ADP}).$$

Then $H = (V, E)$ with $E = \{e_{\mathrm{HK}}, e_{\mathrm{PFK}}\} \subseteq \mathcal{P}(V) \setminus \{\varnothing\}$ is a hypergraph in the sense of Definition 3: each hyperedge joins an arbitrary nonempty subset of metabolites (substrates and products) participating in a reaction.

Definition 4 (n-SuperHyperGraph (cf. [11,42]))**.** Let V_0 be a finite, nonempty *base set*. Define iterated power sets by

$$\mathcal{P}^0(X) := X, \qquad \mathcal{P}^{k+1}(X) := \mathcal{P}\left(\mathcal{P}^k(X)\right) \quad (k \geq 0).$$

Fix $n \geq 1$. An *n-SuperHyperGraph* is a pair

$$\mathrm{SHG}^{(n)} = (V, E),$$

where the *n-supervertex set* satisfies $\varnothing \neq V \subseteq \mathcal{P}^n(V_0)$, and the *n-superedge set* is a nonempty family of nonempty subsets of V,

$$\varnothing \neq E \subseteq \mathcal{P}(V) \setminus \{\varnothing\}.$$

Equivalently, each $e \in E$ is a finite nonempty set of n-supervertices. For $n = 0$, the construction reduces to a standard hypergraph on V_0.

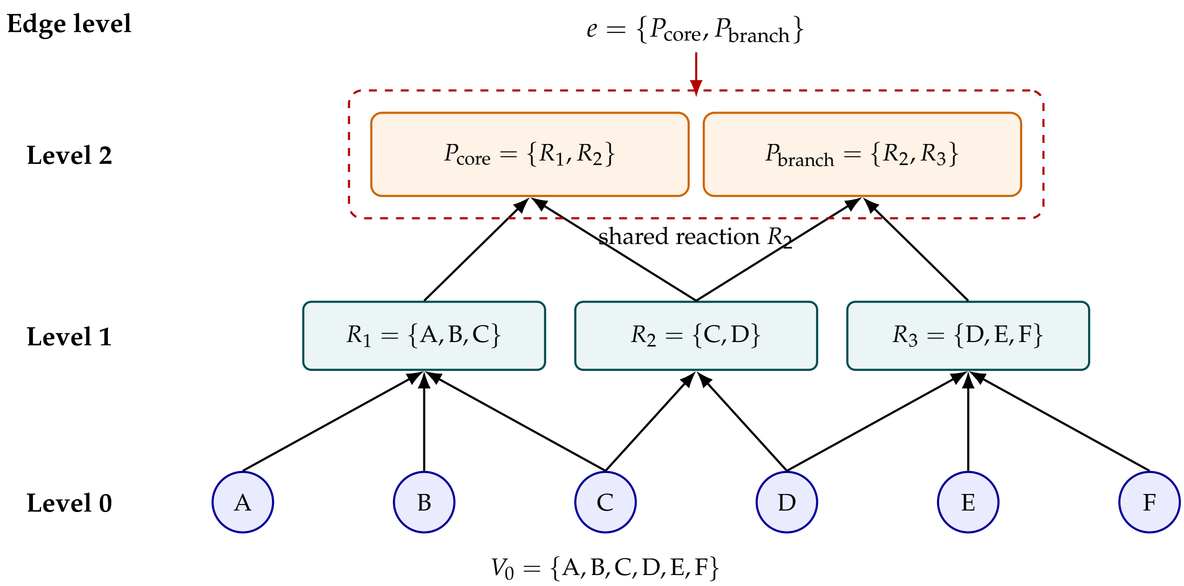


Figure 1. A schematic TikZ illustration of the 2-SuperHyperGraph $\mathrm{SHG}^{(2)} = (V, E)$ in Example 3. The supervertices are the pathways P_{core} and P_{branch}, and the superedge e expresses their crosstalk through the shared reaction R_2.

Example 3 (n=2 SuperHyperGraph: Pathways as Sets of Reactions)**.** Take a finite, nonempty base set V_0 of metabolites,

$$V_0 = \{\mathrm{A, B, C, D, E, F}\}.$$

Consider three reactions modeled as nonempty subsets of metabolites (elements of $\mathcal{P}^1(V_0)$):

$$R_1 = \{\mathrm{A, B, C}\}, \quad R_2 = \{\mathrm{C, D}\}, \quad R_3 = \{\mathrm{D, E, F}\}.$$

Form two metabolic *pathways* as sets of reactions, hence elements of $\mathcal{P}^2(V_0)$:

$$P_{\text{core}} = \{R_1, R_2\}, \qquad P_{\text{branch}} = \{R_2, R_3\}.$$

Define the 2-SuperHyperGraph $\text{SHG}^{(2)} = (V, E)$ by

$$V = \{P_{\text{core}}, P_{\text{branch}}\} \subseteq \mathcal{P}^2(V_0), \qquad E = \{\ \{P_{\text{core}}, P_{\text{branch}}\}\ \} \subseteq \mathcal{P}(V) \setminus \{\varnothing\}.$$

Here, supervertices are pathways (sets of reactions), and the superedge indicates their crosstalk via the shared reaction R_2. This construction satisfies Definition 4 with $n = 2$: $V \subseteq \mathcal{P}^2(V_0)$ and $E \subseteq \mathcal{P}(V) \setminus \{\varnothing\}$.

2.3. Molecular hypergraph and superhypergraph

A Molecular Graph is a mathematical structure used to represent molecules, in which atoms are modeled as vertices and chemical bonds are modeled as edges [43–45]. A Molecular Hypergraph represents molecules where vertices are atoms and hyperedges denote multi-atom interactions or molecular substructures [32,46,47]. The definition of a Molecular Hypergraph is presented as follows.

Definition 5 (Molecular Hypergraph [48])**.** A *molecular hypergraph* is a node- and hyperedge-labeled hypergraph that models the atomic and bonding structure of a molecule. Formally, a molecular hypergraph

$$H = \left(V_H,\ E_H,\ \ell_V^H,\ \ell_E^H\right),$$

consists of:

- V_H a finite set of *nodes*, each representing a chemical bond;
- $E_H \subseteq \mathcal{P}(V_H)$ a finite set of *hyperedges*, where each hyperedge $e \in E_H$ is a subset of V_H corresponding to all bonds incident to a single atom;
- $\ell_V^H : V_H \to L_V^H$ a *node-labeling* function, assigning to each bond–node its bond type (e.g. single, double, triple);
- $\ell_E^H : E_H \to L_E^H$ a *hyperedge-labeling* function, assigning to each atom–hyperedge its atomic symbol or property (e.g. C, O, H).

This structure thus captures molecules at two hierarchical levels: bonds as nodes and atoms as hyperedges

Example 4 (Molecular Hypergraph: Methane CH_4)**.** We instantiate Definition *Molecular Hypergraph* for methane. Let the node set V_H consist of the four C–H bonds:

$$V_H = \{\ b_{CH_1},\ b_{CH_2},\ b_{CH_3},\ b_{CH_4}\ \}.$$

Each hyperedge aggregates all bonds incident to a given atom; thus

$$e_{\text{C}} = \{\ b_{CH_1}, b_{CH_2}, b_{CH_3}, b_{CH_4}\ \},$$
$$e_{H_i} = \{\ b_{CH_i}\ \} \qquad (i = 1, 2, 3, 4).$$

Hence the hyperedge set is

$$E_H = \{\ e_{\text{C}},\ e_{H_1},\ e_{H_2},\ e_{H_3},\ e_{H_4}\ \} \subseteq \mathcal{P}(V_H).$$

Label each bond–node by its bond order via $\ell_V^H(b_{CH_i}) = \text{single}$, and label each atom–hyperedge by its atomic symbol via $\ell_E^H(e_{\text{C}}) = \text{C}$, $\ell_E^H(e_{H_i}) = \text{H}$. This hypergraph represents bonds as nodes and atoms as hyperedges in accordance with the definition.

A molecular n-superhypergraph encodes hierarchical molecular structure by nesting sets of primitive identifiers via iterated powersets up to depth n.

Definition 6 (Molecular n-SuperHypergraph [48])**.** Fix $n \in \mathbb{N}$. Let B be a finite *base set* of primitive molecular identifiers (e.g., bonds or atom–bond pairs). For $k \geq 0$ define the iterated powersets

$$\mathcal{P}^0(B) := B, \qquad \mathcal{P}^{k+1}(B) := \mathcal{P}\left(\mathcal{P}^k(B)\right),$$

and write $\mathcal{P}^{\leq n}(B) := \bigcup_{k=0}^{n} \mathcal{P}^k(B)$. For a set X let $\mathcal{P}^+(X) := \mathcal{P}(X) \setminus \{\varnothing\}$. Define the *rank* (or *tier*) of $x \in \mathcal{P}^{\leq n}(B)$ by

$$\mathrm{rk}(x) := \min\{\, k \in \{0, \dots, n\} : \; x \in \mathcal{P}^k(B) \,\}.$$

A *molecular n-superhypergraph* is a quadruple

$$H = (V, E, \ell_V, \ell_E),$$

consisting of

(i) a vertex set $V \subseteq \mathcal{P}^{\leq n}(B)$;
(ii) an edge set $E \subseteq \mathcal{P}^{\leq n}(V) \cap \mathcal{P}^+(V)$;
(iii) a vertex-labelling map $\ell_V : V \to L_V$;
(iv) an edge-labelling map $\ell_E : E \to L_E$,

subject to the *hierarchy constraint*

$$u \in e \in E \implies \mathrm{rk}(u) < \mathrm{rk}(e).$$

Thus every edge has strictly higher rank than each of its incident members, and all objects in $V \cup E$ have rank at most n. Elements of V ("n-supernodes") may represent nested collections of primitive molecular features (atoms/bonds/groups), while elements of E ("n-superedges") are nested collections of vertices linking such features. For $n = 0$ one recovers a molecular hypergraph $(V \subseteq B,\ E \subseteq \mathcal{P}^+(V))$; for $n = 1$ this yields the usual molecular superhypergraph with one level of nesting.

Example 5 (Molecular 2–SuperHyperGraph: Ethanol CH_3CH_2OH)**.** *Base bonds.* Let

$$B = \{b_{C_1C_2},\ b_{C_1H_{1a}},\ b_{C_1H_{1b}},\ b_{C_1H_{1c}}, b_{C_2H_{2a}},\ b_{C_2H_{2b}},\ b_{C_2O},\ b_{OH}\}.$$

Vertices (atoms as sets of bonds; rank 1*).*

$$\begin{aligned} v_{C_1} &= \{b_{C_1C_2}, b_{C_1H_{1a}}, b_{C_1H_{1b}}, b_{C_1H_{1c}}\}, \\ v_{C_2} &= \{b_{C_1C_2}, b_{C_2H_{2a}}, b_{C_2H_{2b}}, b_{C_2O}\}, \\ v_O &= \{b_{C_2O}, b_{OH}\}, \\ v_{H_{1a}} &= \{b_{C_1H_{1a}}\}, \quad v_{H_{1b}} = \{b_{C_1H_{1b}}\}, \quad v_{H_{1c}} = \{b_{C_1H_{1c}}\}, \\ v_{H_{2a}} &= \{b_{C_2H_{2a}}\}, \quad v_{H_{2b}} = \{b_{C_2H_{2b}}\}, \quad v_{H_O} = \{b_{OH}\}. \end{aligned}$$

Set

$$V = \{v_{C_1}, v_{C_2}, v_O, v_{H_{1a}}, v_{H_{1b}}, v_{H_{1c}}, v_{H_{2a}}, v_{H_{2b}}, v_{H_O}\}.$$

Superedges (groups of atoms; rank 2*).*

$$\begin{aligned} e_{\mathrm{CH_3}} &= \{v_{C_1}, v_{H_{1a}}, v_{H_{1b}}, v_{H_{1c}}\}, \\ e_{\mathrm{CH_2}} &= \{v_{C_2}, v_{H_{2a}}, v_{H_{2b}}\}, \\ e_{\mathrm{OH}} &= \{v_O, v_{H_O}\}. \end{aligned}$$

Thus,

$$E = \{e_{\mathrm{CH_3}}, e_{\mathrm{CH_2}}, e_{\mathrm{OH}}\}.$$

Labels may be chosen as

$$\ell_V(v_{C_1}) = \mathrm{C}, \quad \ell_V(v_{C_2}) = \mathrm{C}, \quad \ell_V(v_O) = \mathrm{O}, \quad \ell_V(v_{H_*}) = \mathrm{H},$$

and

$$\ell_E(e_{CH_3}) = \text{"methyl"}, \quad \ell_E(e_{CH_2}) = \text{"methylene"}, \quad \ell_E(e_{OH}) = \text{"hydroxyl"}.$$

Each superedge $e \in E$ is a nonempty subset of V, so e has rank 2, while each vertex in V has rank 1. Hence this is a molecular 2-SuperHyperGraph.

Example 6 (Molecular 2–SuperHyperGraph: Formaldehyde CH_2O)**.** *Base bonds.*

$$B = \{b_{CO}\ (\text{C=O}),\ b_{CH_1},\ b_{CH_2}\}.$$

Vertices (atoms; rank 1*).*

$$v_C = \{b_{CO}, b_{CH_1}, b_{CH_2}\}, \qquad v_O = \{b_{CO}\},$$
$$v_{H_1} = \{b_{CH_1}\}, \qquad v_{H_2} = \{b_{CH_2}\}.$$

Let

$$V = \{v_C, v_O, v_{H_1}, v_{H_2}\}.$$

Superedges (groups of atoms; rank 2*).*

$$e_{\text{methylene}} = \{v_C, v_{H_1}, v_{H_2}\}, \qquad e_{\text{carbonyl}} = \{v_C, v_O\}.$$

Thus,

$$E = \{e_{\text{methylene}}, e_{\text{carbonyl}}\}.$$

Label maps may be chosen as

$$\ell_V(v_C) = \text{C}, \qquad \ell_V(v_O) = \text{O}, \qquad \ell_V(v_{H_i}) = \text{H},$$

and

$$\ell_E(e_{\text{methylene}}) = \text{"methylene"}, \qquad \ell_E(e_{\text{carbonyl}}) = \text{"carbonyl"}.$$

Each superedge $e \in E$ is a nonempty subset of V, so each such e has rank 2, whereas the vertices in V have rank 1. Therefore, the hierarchy constraint is satisfied, and the structure is a molecular 2-SuperHyperGraph. Figure 2 presents a diagram of the molecular 2-SuperHyperGraph for formaldehyde CH_2O.

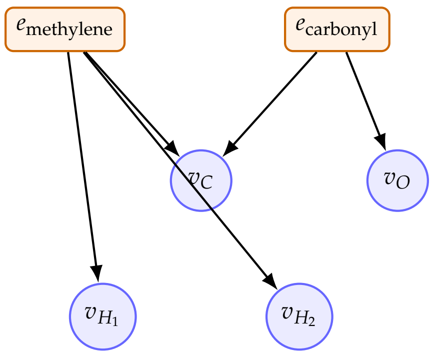


Figure 2. A concise diagram of the molecular 2-SuperHyperGraph for formaldehyde CH_2O

3. Main results

In this section, we present the main results of this paper.

3.1. Weighted molecular SuperHyperGraph

In this subsection, we discuss the Weighted Molecular SuperHyperGraph.

3.1.1. Weighted molecular graph

A Weighted Molecular Graph is a molecular graph where vertices and edges represent atoms and bonds, respectively, each assigned numerical weights capturing chemical properties (cf. [49–51]).

Definition 7 (Weighted Molecular Graph)**.** A *weighted molecular graph* is a finite simple undirected graph

$$G = (V, E),$$

equipped with a *weighting scheme* w that assigns

$$\text{(vertex weights)} \quad V_w : V \to \mathbb{R}_{\geq 0}, \qquad \text{(edge weights)} \quad E_w : E \to \mathbb{R}_{>0},$$

and (optionally) labels

$$\text{(atom types) Sy} : V \to \Sigma, \qquad \text{(bond orders) Bo} : E \to \{1, 1.5, 2, 3\}.$$

In chemical graph theory one writes $G = G(V, E, \text{Sy}, \text{Bo}, V_w, E_w, w)$; vertex/edge weights are computed from atomic properties and bond orders specified by w, which allows one to distinguish heteroatoms and multiple bonds.

For a path $p = v_0 v_1 \cdots v_k$ in G, its w-length is

$$\ell_w(p) = \sum_{t=0}^{k-1} E_w(\{v_t, v_{t+1}\}),$$

and the *weighted distance* between $i, j \in V$ is the minimum path length

$$d_w(i, j) = \min\{\ell_w(p) : p \text{ connects } i \text{ to } j\}.$$

(Topological distance is recovered when all $E_w \equiv 1$.)

Standard weighted molecular matrices (e.g., adjacency $A(w)$, distance $D(w)$, reciprocal distance $RD(w)$) are formed from V_w, E_w, and d_w in the usual way.

Example 7 (Chemistry: Benzene C_6H_6 as a Weighted Molecular Graph)**.** Let $G = (V, E)$ with atoms

$$V = \{C_1, \ldots, C_6,\ H_1, \ldots, H_6\},$$

and edges

$$E = \{\{C_i, C_{i+1}\},\ \{C_i, H_i\}\ :\ i = 1, \ldots, 6,\ C_7 \equiv C_1\}.$$

Assign vertex weights (e.g. atomic numbers) $V_w(C) = 12$, $V_w(H) = 1$. Assign edge weights by bond order: for aromatic C–C bonds set $E_w(\{C_i, C_{i+1}\}) = 1.5$, and for C–H bonds set $E_w(\{C_i, H_i\}) = 1.0$.

For the path from C_1 to H_3,

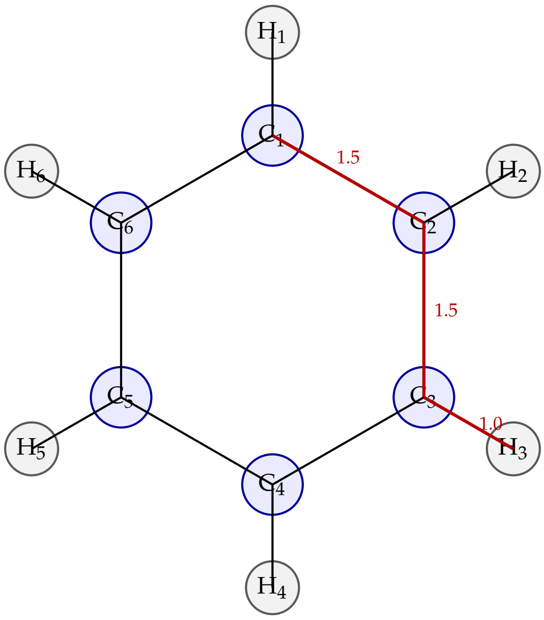
$$p:\ C_1 - C_2 - C_3 - H_3,$$

the w-length is

$$\ell_w(p) = E_w(\{C_1, C_2\}) + E_w(\{C_2, C_3\}) + E_w(\{C_3, H_3\}) = 1.5 + 1.5 + 1.0 = 4.0,$$

and hence $d_w(C_1, H_3) = 4.0$ since any other route uses at least two aromatic C–C edges and one C–H edge as well.

Figure 3 presents a concise diagram of benzene as a weighted molecular graph, with the path $C_1 - C_2 - C_3 - H_3$ highlighted.



C-vertex weight = 12, H-vertex weight = 1, aromatic C–C edge weight = 1.5, C–H edge weight = 1.0

Figure 3. A concise diagram of benzene as a weighted molecular graph. The path $C_1 - C_2 - C_3 - H_3$ is highlighted

3.1.2. Weighted molecular HyperGraph

A Weighted Molecular HyperGraph is a molecular hypergraph in which atom-hyperedges and bond-nodes carry labels and weights, thereby modeling multi-atom interactions with quantitative structural descriptors.

Definition 8 (Weighted Molecular HyperGraph)**.** A *weighted molecular hypergraph* is a tuple

$$\mathcal{H} = \left(V_H, E_H, \iota, \ell_V^H, \ell_E^H, w_V^H, w_E^H\right),$$

where

- V_H is a finite set of *bond-nodes*;
- E_H is a finite set of *atom-hyperedges*;
- $\iota : E_H \to \mathcal{P}(V_H) \setminus \{\varnothing\}$ is the *incidence map*, so each $e \in E_H$ is realized as the nonempty set $\iota(e) \subseteq V_H$ of bonds incident to that atom;
- $\ell_V^H : V_H \to L_V^H$ labels bond-nodes (e.g. order: single/double/triple);
- $\ell_E^H : E_H \to L_E^H$ labels atom-hyperedges (e.g. symbol: C, O, H);
- $w_V^H : V_H \to \mathbb{R}_{\geq 0}$ assigns *bond weights*;
- $w_E^H : E_H \to \mathbb{R}_{>0}$ assigns *atom weights*.

An *alternating hypergraph path* from x to y is a sequence

$$x = v_0,\ e_1,\ v_1,\ e_2,\ \ldots,\ e_k,\ v_k = y,$$

with $v_{t-1}, v_t \in \iota(e_t)$ for $t = 1, \ldots, k$. Its (edge-)weighted length is

$$\ell_w^{\mathrm{H}}(v_0, e_1, v_1, \ldots, e_k, v_k) = \sum_{t=1}^{k} w_E^H(e_t).$$

The corresponding distance $d_w^{\mathrm{H}}(x, y)$ is the minimum such length over all alternating paths from x to y.

Remark 1 (Two equivalent presentations)**.** There are two useful (incidence-dual) presentations:

1. *Bond-centric* (used above): vertices are bonds, hyperedges are atoms collecting incident bonds.
2. *Atom-centric* (useful for reductions to graphs): vertices are atoms, hyperedges are (multi-atom) interactions; when every hyperedge has size 2, this coincides with an ordinary graph.

Passing between the two is achieved by the standard incidence duality on the bipartite incidence graph; we use each where it simplifies proofs.

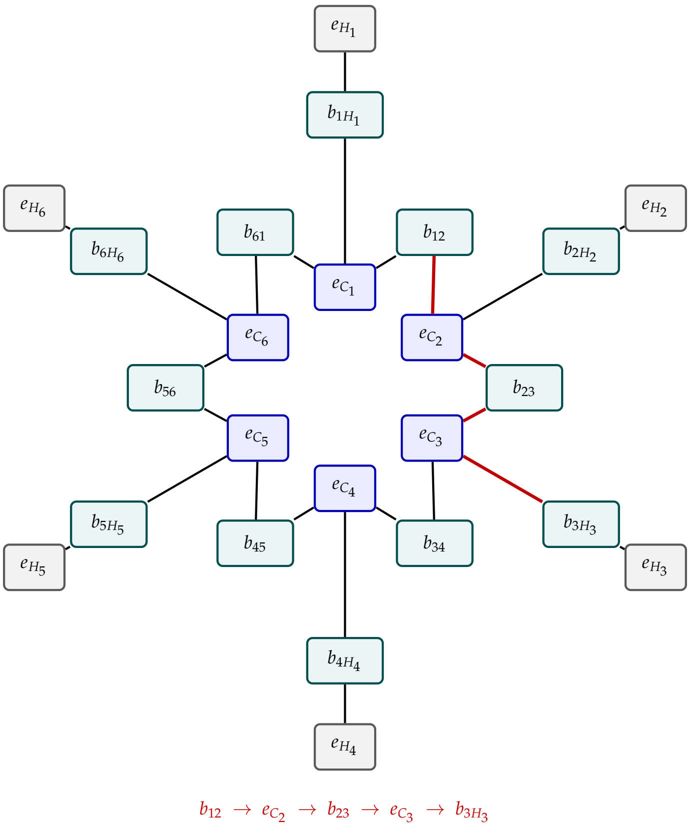


Figure 4. A concise TikZ diagram of the benzene weighted molecular hypergraph (WMHG). Here $b_{12} = b_{C_1C_2}$, . . ., $b_{61} = b_{C_6C_1}$, and $b_{iH_i} = b_{C_iH_i}$

Example 8 (Aromatic hydrocarbon: benzene C_6H_6 as a WMHG)**.** Label the carbon atoms $C_1, \ldots, C_6$ cyclically and the hydrogens $H_1, \ldots, H_6$ with H_i attached to C_i. Define the bond–node set

$$V_H \;=\; \left\{ b_{C_iC_{i+1}} \ (i = 1, \ldots, 6;\ C_7 \equiv C_1),\ \ b_{C_iH_i} \ (i = 1, \ldots, 6) \right\}.$$

Define the atom–hyperedges $E_H = \{e_{C_i} : i = 1, \ldots, 6\} \cup \{e_{H_i} : i = 1, \ldots, 6\}$ with incidence

$$\iota(e_{C_i}) = \left\{ b_{C_{i-1}C_i},\ b_{C_iC_{i+1}},\ b_{C_iH_i} \right\}, \qquad \iota(e_{H_i}) = \left\{ b_{C_iH_i} \right\},$$

where indices are taken modulo 6 (so $C_0 \equiv C_6$, $C_7 \equiv C_1$). Choose labels

$$\ell_V^H(b_{C_iC_{i+1}}) = \text{“aromatic C–C”}, \quad \ell_V^H(b_{C_iH_i}) = \text{“C–H”}, \quad \ell_E^H(e_{C_i}) = \mathrm{C}, \quad \ell_E^H(e_{H_i}) = \mathrm{H}.$$

A chemically meaningful weight scheme is, for instance,

$$w_V^H(b_{C_iC_{i+1}}) = 1.5, \qquad w_V^H(b_{C_iH_i}) = 1.0, \qquad w_E^H(e_{C_i}) = 1.00, \qquad w_E^H(e_{H_i}) = 0.50.$$

Here the vertex weights reflect stronger (delocalized) C–C bonds (1.5) relative to C–H (1.0), while hyperedge weights encode a per–atom traversal cost (C heavier than H).

To illustrate the alternating-path metric, consider the path from $b_{C_1C_2}$ to $b_{C_3H_3}$:

$$b_{C_1C_2} \xrightarrow{e_{C_2}} b_{C_2C_3} \xrightarrow{e_{C_3}} b_{C_3H_3}.$$

Its weighted length (Definition 8) is the sum of the traversed atom–weights:

$$\ell_w^{\mathrm{H}}\left(b_{C_1C_2}, e_{C_2}, b_{C_2C_3}, e_{C_3}, b_{C_3H_3}\right) = w_E^H(e_{C_2}) + w_E^H(e_{C_3}) = 1.00 + 1.00 = 2.00.$$

Any other alternating path between these two bonds must pass through at least two carbon hyperedges, so the shortest distance equals 2.00.

Figure 4 presents a concise TikZ diagram of the benzene weighted molecular hypergraph (WMHG), where $b_{12} = b_{C_1C_2}, \ldots, b_{61} = b_{C_6C_1}$, and $b_{iH_i} = b_{C_iH_i}$.

Lemma 1 (Unweighted molecular hypergraphs embed into WMHGs). *Let $\mathcal{H}_0 = (V_H, E_H, \iota, \ell_V^H, \ell_E^H)$ be any (unweighted) molecular hypergraph in the bond-centric sense. Define*

$$w_V^H(v) \equiv 1 \quad (v \in V_H), \qquad w_E^H(e) \equiv 1 \quad (e \in E_H).$$

Then $\mathcal{H} = (V_H, E_H, \iota, \ell_V^H, \ell_E^H, w_V^H, w_E^H)$ is a WMHG. Moreover, the forgetful map

$$\mathsf{U}: \ \mathcal{H} \mapsto (V_H, E_H, \iota, \ell_V^H, \ell_E^H),$$

satisfies $\mathsf{U}(\mathcal{H}) = \mathcal{H}_0$, i.e. $\mathcal{H}_0$ is recovered exactly by forgetting weights.

Proof. By construction w_V^H and w_E^H are well-defined with codomains $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_{>0}$ (the constant 1 is > 0), so Definition 8 is satisfied. The equalities

$$\mathsf{U}\left(V_H, E_H, \iota, \ell_V^H, \ell_E^H, w_V^H \equiv 1, w_E^H \equiv 1\right) = (V_H, E_H, \iota, \ell_V^H, \ell_E^H),$$

hold componentwise, proving the claim. □

Lemma 2 (Atom-centric WMHG of a weighted graph). *Let $\mathcal{G} = (V_A, E_B, \lambda_V, \lambda_E, w_V, w_E)$ be a Weighted Molecular Graph as in the Definition. Define a WMHG in the atom-centric presentation by*

$$\Phi(\mathcal{G}) \ = \ \left(V_H, E_H, \iota, \tilde{\lambda}_V, \tilde{\lambda}_E, \tilde{w}_V, \tilde{w}_E\right),$$

with

$$V_H := V_A, \qquad E_H := \{\, e_{uv} : \ \{u,v\} \in E_B \,\}, \qquad \iota(e_{uv}) = \{u, v\},$$

and

$$\tilde{\lambda}_V := \lambda_V, \quad \tilde{\lambda}_E(e_{uv}) := \lambda_E(\{u,v\}), \qquad \tilde{w}_V := w_V, \quad \tilde{w}_E(e_{uv}) := w_E(\{u,v\}).$$

Then $\Phi(\mathcal{G})$ is a WMHG and for every graph path $p = v_0 v_1 \cdots v_k$ one has the alternating WMHG path

$$\Pi(p) \ = \ v_0, \ e_{v_0v_1}, \ v_1, \ e_{v_1v_2}, \ \ldots, \ e_{v_{k-1}v_k}, \ v_k,$$

with identical length:

$$\ell_w(p) = \sum_{t=1}^{k} w_E(\{v_{t-1}, v_t\}) = \sum_{t=1}^{k} \tilde{w}_E(e_{v_{t-1}v_t}) = \ell_{\tilde{w}}^{\mathrm{H}}(\Pi(p)).$$

Consequently,

$$d_w(i,j) \ = \ d_{\tilde{w}}^{\mathrm{H}}(i,j) \qquad \text{for all } i, j \in V_A (= V_H).$$

Proof. By construction $|\iota(e_{uv})| = 2$, hence $\Phi(\mathcal{G})$ satisfies Definition 8. The displayed equality follows term-by-term from the definitions $\tilde{w}_E(e_{uv}) = w_E(\{u,v\})$ and the fact that the alternating path $\Pi(p)$ uses exactly the hyperedges $e_{v_{t-1}v_t}$ in the same order. Taking minima over all paths on both sides yields $d_w = d^{\mathrm{H}}_{\tilde{w}}$. □

Theorem 1 (WMHG strictly generalizes WMG and MH)**.** *The class of Weighted Molecular HyperGraphs contains:*

(1) *every Molecular HyperGraph (via Lemma 1 with constant weights), and*

(2) *every Weighted Molecular Graph (via the atom-centric embedding Φ of Lemma 2).*

Moreover, in case (2) the edge-weighted shortest-path metric of the graph is preserved verbatim by the WMHG alternating-path metric.

Proof. Item (1) is immediate from Lemma 1. For (2), Lemma 2 furnishes an injective construction Φ whose incidence sets are size-2 hyperedges, with $\tilde{w}_E = w_E$ on corresponding edges. The length identity

$$\ell_w(p) = \ell^{\mathrm{H}}_{\tilde{w}}\left(\Pi(p)\right),$$

holds for every path p, hence the minima coincide, $d_w = d^{\mathrm{H}}_{\tilde{w}}$. This proves preservation of the metric and hence the generalization claim. □

3.1.3. Weighted molecular SuperHyperGraph

We introduce a rank–aware, weighted superhypergraph model for molecules that subsumes both the (unweighted) Molecular SuperHyperGraph and the Weighted Molecular HyperGraph.

Notation 1. Let B be a finite *base set* of primitive molecular identifiers (e.g., bonds or atom–bond pairs). Define the iterated power sets $\mathcal{P}^0(X) := X$ and $\mathcal{P}^{k+1}(X) := \mathcal{P}\left(\mathcal{P}^k(X)\right)$ for $k \geq 0$, and the truncated union

$$\mathcal{P}^{\leq n}(B) := \bigcup_{k=0}^{n} \mathcal{P}^k(B), \qquad n \in \mathbb{N}.$$

For $x \in \mathcal{P}^{\leq n}(B)$, its *rank* (or *tier*) is

$$\mathrm{rk}(x) := \min\{\, k \in \{0, \dots, n\} : x \in \mathcal{P}^k(B)\,\}.$$

Definition 9 (Weighted Molecular SuperHyperGraph (WMSHG))**.** Fix $n \in \mathbb{N}$. A *weighted molecular n–superhypergraph* is a tuple

$$\mathfrak{H}^{(n)} = (V,\ E,\ \ell_V,\ \ell_E,\ w_V,\ w_E),$$

consisting of

(i) a nonempty *vertex set* $V \subseteq \mathcal{P}^{\leq n}(B)$;
(ii) a nonempty *edge (superedge) set* $E \subseteq \mathcal{P}^{\leq n}(V) \cap \mathcal{P}(V) \setminus \{\varnothing\}$;
(iii) label maps $\ell_V : V \to L_V$ and $\ell_E : E \to L_E$;
(iv) weight maps

$$w_V : V \to \mathbb{R}_{\geq 0}, \qquad w_E : E \to \mathbb{R}_{>0},$$

subject to the *hierarchy constraint*

$$u \in e \in E \quad \Longrightarrow \quad \mathrm{rk}(u) < \mathrm{rk}(e).$$

An *alternating superhypergraph path* is a finite sequence

$$p = v_0,\ e_1,\ v_1,\ e_2,\ \dots,\ e_m,\ v_m,$$

with $v_{t-1}, v_t \in e_t$ for $t = 1, \dots, m$. Its (edge–)weighted length is

$$\ell^{\mathrm{SH}}_w(p) := \sum_{t=1}^{m} w_E(e_t).$$

The associated distance on V is

$$d_w^{\mathrm{SH}}(x,y) \;:=\; \min\Big\{\, \ell_w^{\mathrm{SH}}(p) \;:\; p \text{ alternating path from } x \text{ to } y \,\Big\}.$$

Remark 2. (i) Only edge weights w_E contribute to ℓ_w^{SH}, matching the molecular–hypergraph convention that traversal cost accrues at atoms/superatoms; w_V is available for descriptors or alternative metrics. (ii) The hierarchy constraint enforces that every superedge has strictly higher rank than its incident vertices, so membership is always "upward" in rank.

Example 9 (Ethanol–water hydrogen-bonded microcomplex as a WMSHG)**.** We build a concrete *Weighted Molecular SuperHyperGraph* (WMSHG) of depth $n = 2$ for liquid–phase ethanol CH_3CH_2OH solvated by one water molecule, capturing functional-group structure and an O–H$\cdots$O hydrogen bond to water.

3.1.4. Backbone (MSHG at depth $n = 2$)

Let B be the set of *covalent* bonds:

$$B = \{b_{C_1C_2},\, b_{C_1H_{1a}},\, b_{C_1H_{1b}},\, b_{C_1H_{1c}},\, b_{C_2H_{2a}},\, b_{C_2H_{2b}},\, b_{C_2O},\, b_{OH},\, b_{O_wH_{w1}},\, b_{O_wH_{w2}}\}.$$

Rank-1 vertices (*atoms as sets of bonds*):

$$v_{C_1} = \{b_{C_1C_2}, b_{C_1H_{1a}}, b_{C_1H_{1b}}, b_{C_1H_{1c}}\},\quad v_{C_2} = \{b_{C_1C_2}, b_{C_2H_{2a}}, b_{C_2H_{2b}}, b_{C_2O}\},$$
$$v_O = \{b_{C_2O}, b_{OH}\},\quad v_{H_O} = \{b_{OH}\},\quad v_{H_{1a}} = \{b_{C_1H_{1a}}\},\ v_{H_{1b}} = \{b_{C_1H_{1b}}\},\ v_{H_{1c}} = \{b_{C_1H_{1c}}\},$$
$$v_{H_{2a}} = \{b_{C_2H_{2a}}\},\ v_{H_{2b}} = \{b_{C_2H_{2b}}\},\quad v_{O_w} = \{b_{O_wH_{w1}}, b_{O_wH_{w2}}\},\ v_{H_{w1}} = \{b_{O_wH_{w1}}\},\ v_{H_{w2}} = \{b_{O_wH_{w2}}\}.$$

Set V to be the collection of the above rank-1 vertices.
Rank-2 *superedges* (each is a subset of V; hierarchy $\mathrm{rk}(u) < \mathrm{rk}(e)$ holds):

$$A_{\mathrm{CH_3}} := \{v_{C_1}, v_{H_{1a}}, v_{H_{1b}}, v_{H_{1c}}\},\qquad A_{\text{C-skel}} := \{v_{C_1}, v_{C_2}\},$$
$$A_{\mathrm{CO}} := \{v_{C_2}, v_O\},\qquad A_{\mathrm{OH}} := \{v_O, v_{H_O}\},\qquad A_{\mathrm{H_2O}} := \{v_{O_w}, v_{H_{w1}}, v_{H_{w2}}\},$$
$$e_{\mathrm{HB}} := \{v_O,\ v_{H_{w1}}\}\ \text{(O–H}\cdots\text{O hydrogen bond).}$$

Define $E := \{A_{\mathrm{CH_3}}, A_{\text{C-skel}}, A_{\mathrm{CO}}, A_{\mathrm{OH}}, A_{\mathrm{H_2O}}, e_{\mathrm{HB}}\}$.

3.1.5. Labels and weights

Let ℓ_V map atoms to symbols $\{C, O, H\}$ and ℓ_E map each superedge to its type (e.g. "CH_3 group", "C-skeleton", "C–O link", "OH group", "H_2O", "H-bond"). Choose *vertex* weights (for descriptors) $w_V : V \to \mathbb{R}_{\geq 0}$, e.g.

$$w_V(v_{C_\bullet}) = 12.0,\quad w_V(v_O) = 16.0,\ w_V(v_{O_w}) = 16.0,\quad w_V(v_{H_\bullet}) = 1.0,$$

and *edge* weights (used in path length) $w_E : E \to \mathbb{R}_{>0}$:

$$w_E(A_{\mathrm{CH_3}}) = 0.60,\ \ w_E(A_{\text{C-skel}}) = 0.80,\ \ w_E(A_{\mathrm{CO}}) = 1.10,$$

$$w_E(A_{\mathrm{OH}}) = 0.70,\ \ w_E(A_{\mathrm{H_2O}}) = 0.90,\ \ w_E(e_{\mathrm{HB}}) = 0.35.$$

(Heavier weights reflect stronger/less flexible intramolecular relations; the hydrogen bond is lighter.)

3.1.6. Alternating-path distance (explicit calculation)

Consider the (rank-1) vertices $s := v_{H_{1a}}$ (a methyl hydrogen) and $t := v_{H_{w1}}$ (a water hydrogen). One alternating path is

$$p:\quad v_{H_{1a}} \xrightarrow{A_{\mathrm{CH_3}}} v_{C_1} \xrightarrow{A_{\text{C-skel}}} v_{C_2} \xrightarrow{A_{\mathrm{CO}}} v_O \xrightarrow{e_{\mathrm{HB}}} v_{H_{w1}}.$$

Its edge–weighted length is

$$\ell_w^{SH}(p) = w_E(A_{CH_3}) + w_E(A_{C\text{-skel}}) + w_E(A_{CO}) + w_E(e_{HB}) = 0.60 + 0.80 + 1.10 + 0.35 = 2.85.$$

3.1.7. Minimality (shortest-path certificate).

Define the following nested vertex separators

$$S_0 := \{v_{H_{1a}}, v_{H_{1b}}, v_{H_{1c}}\}, \quad S_1 := \{v_{C_1}\}, \quad S_2 := \{v_{C_2}\}, \quad S_3 := \{v_O\}, \quad S_4 := \{v_{H_{w1}}\}.$$

By construction of E, the *only* superedges intersecting consecutive pairs (S_0, S_1), (S_1, S_2), (S_2, S_3), (S_3, S_4) are, respectively, A_{CH_3}, $A_{C\text{-skel}}$, A_{CO}, and e_{HB}, and there is *no* superedge that skips any separator (e.g. none connects S_1 directly to S_3). Hence any alternating path from $s \in S_0$ to $t \in S_4$ must contain at least one edge from each of these four cut-sets; since each cut-set is a singleton, every such path uses exactly the four edges above. Because all edge weights are strictly positive, every s–t path has length at least $w_E(A_{CH_3}) + w_E(A_{C\text{-skel}}) + w_E(A_{CO}) + w_E(e_{HB})$, which is achieved by p. Therefore

$$d_w^{SH}(v_{H_{1a}}, v_{H_{w1}}) = 2.85.$$

The WMSHG encodes a multi-level, *weighted* molecular architecture: motion from a methyl proton to a water proton must traverse (i) the methyl group, (ii) the carbon skeleton, (iii) the C–O connection, and (iv) the intermolecular hydrogen bond. The learned or designed weights w_E can reflect energetic/entropic costs, enabling property prediction or pathway analysis on the superhypergraph. The graph corresponding to this concrete example is shown in Figure 5.

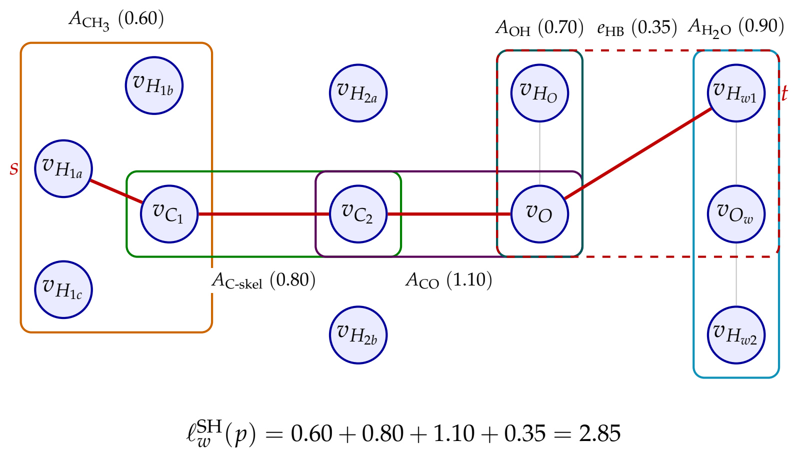


Figure 5. A concise diagram of the weighted molecular superhypergraph (WMSHG) for the ethanol–water hydrogen-bonded microcomplex. Rank-1 vertices are atoms, while rank-2 superedges are shown as fitted regions. The highlighted path from $v_{H_{1a}}$ to $v_{H_{w1}}$ realizes distance 2.85

Lemma 3 (Unweighted MSHG is a special case of WMSHG). *Let* $\mathfrak{H}_0^{(n)} = (V, E, \ell_V, \ell_E)$ *be any (unweighted) molecular n–superhypergraph satisfying the hierarchy constraint. Define constant weights*

$$w_V(v) \equiv 0 \quad (v \in V), \qquad w_E(e) \equiv 1 \quad (e \in E).$$

Then $\mathfrak{H}^{(n)} = (V, E, \ell_V, \ell_E, w_V, w_E)$ *is a WMSHG. Moreover, the forgetful map*

$$\mathsf{U} : (V, E, \ell_V, \ell_E, w_V, w_E) \longmapsto (V, E, \ell_V, \ell_E),$$

satisfies $\mathsf{U}(\mathfrak{H}^{(n)}) = \mathfrak{H}_0^{(n)}$, *and the unweighted alternating–path distance (edge count) is recovered as*

$$\#\text{edges along a shortest alternating path from } x \text{ to } y \; = \; d_w^{SH}(x, y) \quad (\text{since } w_E \equiv 1).$$

Proof. By assumption, (V, E, ℓ_V, ℓ_E) already satisfies the hierarchy constraint and $E \subseteq \mathcal{P}(V) \setminus \{\varnothing\}$. The choices $w_V \equiv 0$ and $w_E \equiv 1$ meet the codomain requirements $w_V : V \to \mathbb{R}_{\geq 0}$ and $w_E : E \to \mathbb{R}_{>0}$, so Definition 9 holds. For any alternating path p with edges $e_1, \ldots, e_m$,

$$\ell_w^{\mathrm{SH}}(p) = \sum_{t=1}^{m} w_E(e_t) = \sum_{t=1}^{m} 1 = m,$$

hence minimizing ℓ_w^{SH} equals minimizing the edge count; this equals the usual unweighted alternating–path distance. The stated equality follows. □

We work with the *bond–centric* Weighted Molecular HyperGraph (WMHG): vertices are bonds, and each atom is a hyperedge collecting its incident bonds.

Lemma 4 (WMHG embeds into n=1 WMSHG)**.** *Let* $\mathcal{H}_{\mathrm{WMHG}} = (V_H, E_H, \iota, \ell_V^H, \ell_E^H, w_V^H, w_E^H)$ *be a WMHG, where* $\iota : E_H \to \mathcal{P}(V_H) \setminus \{\varnothing\}$ *is the incidence map. Construct a* 1*–superhypergraph*

$$\Phi(\mathcal{H}_{\mathrm{WMHG}}) = (V,\ E,\ \ell_V,\ \ell_E,\ w_V,\ w_E),$$

by setting the base $B := V_H$ *and*

$$V := V_H \ \ (\text{rank } 0), \qquad E := \{\, \iota(e) \ :\ e \in E_H \,\} \subseteq \mathcal{P}(V) \ \ (\text{rank } 1),$$

$$\ell_V := \ell_V^H, \qquad \ell_E\,(\iota(e)) := \ell_E^H(e), \qquad w_V := w_V^H, \qquad w_E\,(\iota(e)) := w_E^H(e).$$

Then $\Phi(\mathcal{H}_{\mathrm{WMHG}})$ *is a WMSHG of depth* $n = 1$ *satisfying the hierarchy constraint, and for any WMHG alternating path*

$$p_H = v_0,\ e_1,\ v_1,\ \ldots,\ e_m,\ v_m \qquad (v_{t-1}, v_t \in \iota(e_t)),$$

the corresponding WMSHG path

$$p_{SH} = v_0,\ \iota(e_1),\ v_1,\ \ldots,\ \iota(e_m),\ v_m,$$

has identical length:

$$\ell_w^{\mathrm{SH}}(p_{SH}) = \sum_{t=1}^{m} w_E\,(\iota(e_t)) = \sum_{t=1}^{m} w_E^H(e_t) = \ell_w^{\mathrm{H}}(p_H).$$

Consequently, the WMHG distance equals the WMSHG distance:

$$d_w^{\mathrm{H}}(x, y) = d_w^{\mathrm{SH}}(x, y) \qquad \text{for all } x, y \in V_H (= V).$$

Proof. By construction, $V \subseteq \mathcal{P}^{\leq 1}(B)$ with $\mathrm{rk}(v) = 0$ for $v \in V = V_H$, and $E \subseteq \mathcal{P}(V)$ with $\mathrm{rk}(e) = 1$ for $e \in E$. Hence whenever $v \in e$ we have $\mathrm{rk}(v) = 0 < 1 = \mathrm{rk}(e)$, i.e., the hierarchy constraint holds, and Definition 9 is satisfied. The length equality is a termwise identity:

$$\ell_w^{\mathrm{SH}}(p_{SH}) = \sum_t w_E\,(\iota(e_t)) = \sum_t w_E^H(e_t) = \ell_w^{\mathrm{H}}(p_H),$$

so minimizing over paths on both sides yields $d_w^{\mathrm{H}} = d_w^{\mathrm{SH}}$. □

Theorem 2 (WMSHG jointly generalizes WMHG and MSHG)**.** *For every unweighted Molecular n–SuperHyperGraph*

$$\mathfrak{H}_0^{(n)},$$

there exists a WMSHG $\mathfrak{H}^{(n)}$ *with*

$$\mathsf{U}(\mathfrak{H}^{(n)}) = \mathfrak{H}_0^{(n)},$$

and identical unweighted alternating–path distances (Lemma 3). For every Weighted Molecular HyperGraph $\mathcal{H}_{\mathrm{WMHG}}$ *there exists a depth–1 WMSHG* $\Phi(\mathcal{H}_{\mathrm{WMHG}})$ *preserving labels, weights, incidence, and alternating–path distances exactly (Lemma 4). Hence the class of WMSHGs strictly contains both the class of MSHGs and the class of WMHGs.*

Proof. The two inclusions are given by Lemma 3 and Lemma 4. Strictness follows because WMSHGs allow simultaneously (a) higher ranks $n \geq 2$ with nested superedges and (b) nontrivial edge weights w_E, while an MSHG lacks weights and a WMHG (in bond–centric form) has depth $n = 1$ only. □

3.2. Rough molecular SuperHyperGraph

In this subsection, we discuss concepts related to the Rough Molecular SuperHyperGraph.

3.2.1. Rough molecular Graph

Rough Set provides a mathematical tool to handle uncertain information by approximating the set of elements [52,53]. Rough Graph extends Rough Set Theory to graphs, where uncertainty in relationships (edges) is represented through lower and upper approximations [54].

Definition 10 (Rough Set)**.** [55–57] Let U be a universe of discourse and R a relation on U that induces a partition of U into equivalence classes. For a subset $X \subseteq U$, the *lower approximation* of X with respect to R (denoted $R(X)$) is the set of elements that are certainly in X given the information provided by R:

$$R(X) = \{x \in U \mid [x]_R \subseteq X\}.$$

The *upper approximation* of X (denoted $R(X)$) is the set of elements that possibly belong to X:

$$R(X) = \{x \in U \mid [x]_R \cap X \neq \emptyset\}.$$

The pair $(R(X), R(X))$ is called the *Rough Set* approximation of X with respect to R.

Definition 11 (Rough Graph [58–60])**.** Let $G = (V, E)$ be a graph where V is the set of vertices and E is the set of edges. Let R be an attribute set on E, inducing an equivalence relation on the edges. For any edge set $X \subseteq E$, the *lower approximation* of X with respect to R (denoted $R(X)$) is defined as:

$$R(X) = \{e \in E \mid [e]_R \subseteq X\},$$

where $[e]_R$ denotes the equivalence class of e under R. The *upper approximation* of X (denoted $R(X)$) is defined as:

$$R(X) = \{e \in E \mid [e]_R \cap X \neq \emptyset\}.$$

A graph $G = (V, E)$ is called an *R-rough graph* if X is not exactly definable under R, and it is characterized by the pair $(R(X), R(X))$, where $R(X)$ is the *lower approximation graph* and $R(X)$ is the *upper approximation graph*.

A Rough Molecular Graph models molecules with indiscernibility relations on atoms and bonds, providing lower and upper approximations of substructures.

Definition 12 (Rough Molecular Graph (RMG))**.** Let $G = (V, E, \lambda_V, \lambda_E)$ be a molecular graph as in the Definition. Fix two equivalence relations

$$\sim_V \quad \text{on } V, \qquad \sim_E \quad \text{on } E,$$

called the *vertex* and *edge indiscernibility relations*. Typically, they are induced by attributes, e.g.

$$u \sim_V v \iff \lambda_V(u) = \lambda_V(v), \qquad e \sim_E f \iff \lambda_E(e) = \lambda_E(f).$$

For $X \subseteq V$ and $Y \subseteq E$, let $\underline{X}, \overline{X}$ be computed from $\sim_V$, and $\underline{Y}, \overline{Y}$ from $\sim_E$ via the Definition.

Given any *crisp target* subgraph $S = (X, Y)$ with $X \subseteq V$ and $Y \subseteq E$ satisfying

$$Y \subseteq E[X] := \{\{i, j\} \in E : i, j \in X\},$$

the *lower* and *upper rough molecular subgraphs* of S are

$$\underline{S} := (\, \underline{X},\, \underline{Y} \cap E[\underline{X}]\,), \qquad \overline{S} := (\, \overline{X},\, \overline{Y} \cap E[\overline{X}]\,).$$

The tuple

$$\mathcal{G}_{\mathrm{RMG}} := \Big(G,\, \sim_V,\, \sim_E,\, \underline{(\cdot)},\, \overline{(\cdot)}\Big),$$

is called a *Rough Molecular Graph*.

Example 10 (Acetic acid in water as a Rough Molecular Graph)**.** We give a real–life RMG where coarse experimental resolution (e.g., fast proton exchange and resonance) makes the two carboxyl oxygens indistinguishable at graph level.

Underlying molecular graph $G = (V, E, \lambda_V, \lambda_E)$.

Let the atoms be

$$V = \{C_1, C_2, O_d, O_s, H_a, H_{1a}, H_{1b}, H_{1c}\},$$

where O_d denotes the carbonyl oxygen, O_s the hydroxyl oxygen, and H_a the acidic proton. The covalent bonds are

$$E = \{\ \{C_1, C_2\},\ \{C_2, O_d\},\ \{C_2, O_s\},\ \{O_s, H_a\},\ \{C_1, H_{1a}\},\ \{C_1, H_{1b}\},\ \{C_1, H_{1c}\}\ \}.$$

Labels encode coarse chemistry:

$$\lambda_V(\cdot) \in \{\mathrm{C}, \mathrm{O}, \mathrm{H}\}, \qquad \lambda_E(\{u, v\}) \in \{\mathrm{C–C}, \mathrm{C–O}, \mathrm{O–H}, \mathrm{C–H}\}.$$

3.2.2. Indiscernibility relations

At typical aqueous conditions, the carboxylate fragment exhibits resonance and fast proton exchange between O_d and O_s. We capture this by

$$O_d \sim_V O_s \quad (\text{“carboxyl oxygens”}), \qquad \{C_2, O_d\} \sim_E \{C_2, O_s\} \quad (\text{“carboxyl C–O bonds”}),$$

and use identity on all other vertices/edges. These $\sim_V, \sim_E$ are admissible per Definition 12 (e.g., induced by coarse labels “carboxyl oxygen” / “carboxyl C–O”).

3.2.3. A crisp target subgraph and its rough approximations

Suppose a snapshot–level detection aims at the *carbonyl fragment only*:

$$X := \{C_2, O_d\} \subseteq V, \qquad Y := \{\{C_2, O_d\}\} \subseteq E.$$

Clearly $Y \subseteq E[X]$ so $S = (X, Y)$ is an admissible crisp subgraph. Compute the rough sets.
Vertices. The $\sim_V$–class of O_d is $[O_d]_{\sim_V} = \{O_d, O_s\} \nsubseteq X$, hence

$$\underline{X} = \{v \in V : [v]_{\sim_V} \subseteq X\} = \varnothing, \qquad \overline{X} = \{v \in V : [v]_{\sim_V} \cap X \neq \varnothing\} = \{C_2, O_d, O_s\}.$$

Edges. The $\sim_E$–class of $\{C_2, O_d\}$ is $[\{C_2, O_d\}]_{\sim_E} = \{\{C_2, O_d\}, \{C_2, O_s\}\} \nsubseteq Y$, thus

$$\underline{Y} = \{e \in E : [e]_{\sim_E} \subseteq Y\} = \varnothing, \qquad \overline{Y} = \{e \in E : [e]_{\sim_E} \cap Y \neq \varnothing\} = \{\{C_2, O_d\}, \{C_2, O_s\}\}.$$

Induced-edge filters. Using $E[U] := \{\{i, j\} \in E :\ i, j \in U\}$,

$$E[\underline{X}] = E[\varnothing] = \varnothing, \qquad E[\overline{X}] = E[\{C_2, O_d, O_s\}] = \{\{C_2, O_d\}, \{C_2, O_s\}\}.$$

Rough subgraphs (Definition 12). Therefore

$$\underline{S} = (\underline{X},\, \underline{Y} \cap E[\underline{X}]) = (\varnothing, \varnothing), \qquad \overline{S} = \big(\overline{X},\, \overline{Y} \cap E[\overline{X}]\big) = (\{C_2, O_d, O_s\},\ \{\{C_2, O_d\}, \{C_2, O_s\}\}).$$

At coarse resolution that cannot distinguish O_d from O_s, the *certain* (lower) subgraph contains no specific carbonyl assignment, while the *possible* (upper) subgraph necessarily includes *both* oxygens and *both* C–O bonds. This reflects the real aqueous behavior of acetic acid/carboxylate, where resonance and rapid proton exchange make the identity of the "carbonyl" oxygen indiscernible in aggregated observations. The graph of this concrete example is shown in Figure 6.

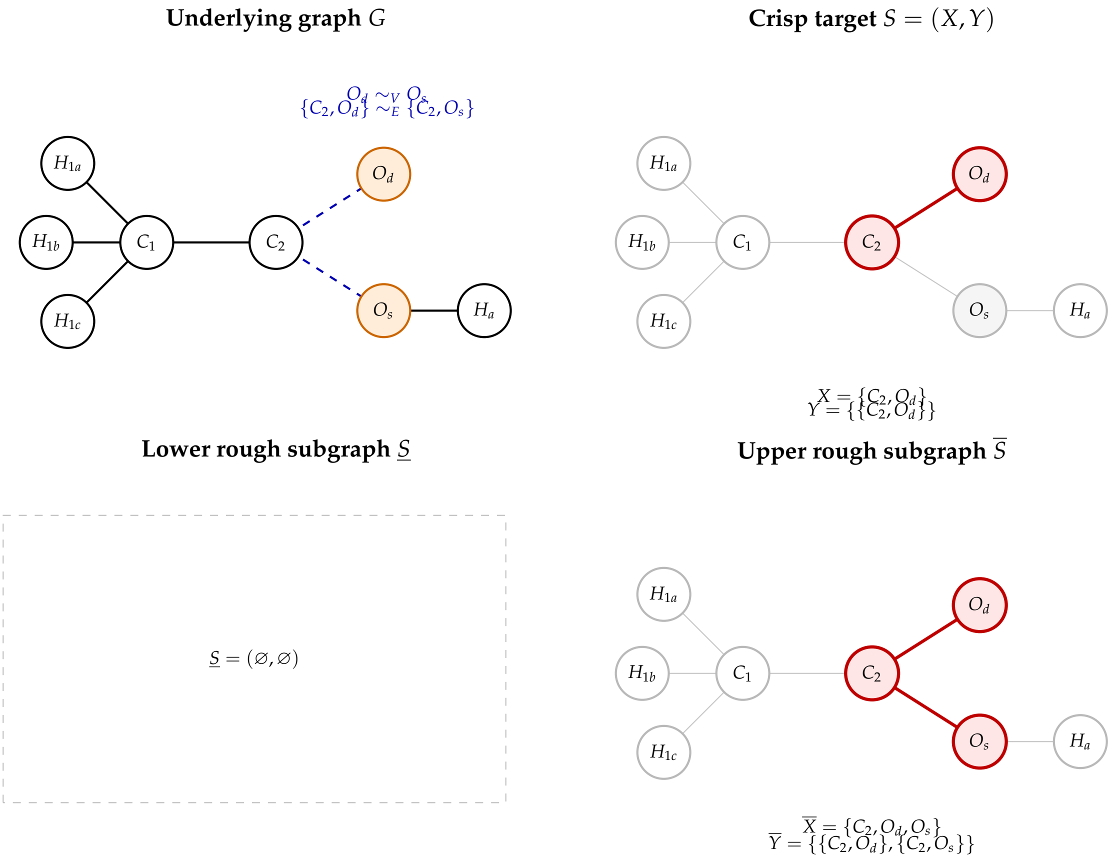


Figure 6. A illustration of the rough molecular graph for acetic acid in water. The two carboxyl oxygens are indiscernible at coarse resolution, so the lower approximation of the carbonyl target is empty, whereas the upper approximation contains both oxygens and both C–O bonds

Lemma 5 (Well-defined rough subgraphs)**.** *For every crisp target subgraph $S = (X, Y)$ of G, the pairs $\underline{S}$ and $\overline{S}$ in Definition 12 are (crisp) subgraphs of G. Moreover,*

$$\underline{S} \subseteq S \subseteq \overline{S},$$

in the sense that vertex and edge sets are included componentwise.

Proof. We verify the lower case; the upper case is analogous. Since $\underline{X} \subseteq X$, every edge of $E[\underline{X}]$ joins two vertices in $\underline{X}$, hence $E[\underline{X}]$ is the induced edge set on $\underline{X}$. Intersecting with $\underline{Y}$ can only remove edges, so $\underline{Y} \cap E[\underline{X}] \subseteq E[\underline{X}]$, i.e. $\underline{S}$ is a subgraph with vertex set $\underline{X}$ and edge set $\underline{Y} \cap E[\underline{X}]$.

Componentwise inclusion holds as follows. For vertices, $\underline{X} \subseteq X$. For edges, we have $\underline{Y} \subseteq Y$ and $E[\underline{X}] \subseteq E[X]$, hence

$$\underline{Y} \cap E[\underline{X}] \subseteq Y \cap E[X] = Y,$$

using the assumption $Y \subseteq E[X]$. Thus $\underline{S} \subseteq S$. The inclusion $S \subseteq \overline{S}$ follows from $X \subseteq \overline{X}$ and $Y \subseteq \overline{Y}$ together with $E[X] \subseteq E[\overline{X}]$. □

Lemma 6 (Monotonicity and idempotence). *Let $S_1 = (X_1, Y_1)$ and $S_2 = (X_2, Y_2)$ be crisp subgraphs of G with $S_1 \subseteq S_2$. Then*

$$\underline{S_1} \subseteq \underline{S_2}, \qquad \overline{S_1} \subseteq \overline{S_2},$$

and the operators are idempotent:

$$\underline{(\underline{S})} = \underline{S}, \qquad \overline{(\overline{S})} = \overline{S}.$$

Proof. From $X_1 \subseteq X_2$ and $Y_1 \subseteq Y_2$, it gives $\underline{X_1} \subseteq \underline{X_2}$ and $\underline{Y_1} \subseteq \underline{Y_2}$, hence

$$\underline{Y_1} \cap E[\underline{X_1}] \subseteq \underline{Y_2} \cap E[\underline{X_2}],$$

because $E[\underline{X_1}] \subseteq E[\underline{X_2}]$ when $\underline{X_1} \subseteq \underline{X_2}$. This yields $\underline{S_1} \subseteq \underline{S_2}$. The upper case is analogous. Idempotence follows from $\underline{\underline{X}} = \underline{X}$, $\underline{\underline{Y}} = \underline{Y}$ and $E[\underline{\underline{X}}] = E[\underline{X}]$; similarly for $\overline{(\cdot)}$. □

Theorem 3 (RMG generalizes the molecular graph). *Every (crisp) molecular graph $G = (V, E, \lambda_V, \lambda_E)$ arises as a special case of a Rough Molecular Graph. In particular, if we take the* discrete *relations*

$$\sim_V = \text{ equality on } V, \qquad \sim_E = \text{ equality on } E,$$

then for every crisp target subgraph $S = (X, Y)$ with $Y \subseteq E[X]$ one has

$$\underline{S} = S = \overline{S}.$$

Hence the functor

$$\mathsf{Emb} : \textit{(Molecular Graphs)} \longrightarrow \textit{(Rough Molecular Graphs)}, \qquad G \longmapsto \Big(G, =_V, =_E, \underline{(\cdot)}, \overline{(\cdot)}\Big),$$

is an injective (full and faithful on objects) embedding of structures.

Proof. Fix $\sim_V$ and $\sim_E$ to be equality on V and E respectively. Then every equivalence class is a singleton:

$$\forall v \in V : [v]_{\sim_V} = \{v\}, \qquad \forall e \in E : [e]_{\sim_E} = \{e\}.$$

By Definition,

$$\underline{X} = \{v \in V : [v]_{\sim_V} \subseteq X\} = \{v \in V : \{v\} \subseteq X\} = X,$$

and

$$\overline{X} = \{v \in V : [v]_{\sim_V} \cap X \neq \varnothing\} = \{v \in V : \{v\} \cap X \neq \varnothing\} = X.$$

The same computation gives $\underline{Y} = Y = \overline{Y}$ for all $Y \subseteq E$. Therefore,

$$\underline{S} = (\underline{X}, \underline{Y} \cap E[\underline{X}]) = (X, Y \cap E[X]) = (X, Y) = S,$$

where the last equality uses the assumption $Y \subseteq E[X]$. Similarly,

$$\overline{S} = \big(\overline{X}, \overline{Y} \cap E[\overline{X}]\big) = (X, Y \cap E[X]) = S.$$

Hence, with discrete relations, the RMG reduces exactly to the original crisp molecular graph, and the assignment $G \mapsto \Big(G, =_V, =_E, \underline{(\cdot)}, \overline{(\cdot)}\Big)$ realizes G as a special case of an RMG. Injectivity on objects is immediate. □

3.2.4. Rough molecular HyperGraph

We lift rough approximations to the molecular *hyper*graph level by allowing indiscernibility on both bonds (hypergraph vertices, in the bond–centric presentation) and atoms (hyperedges). Throughout we work with a fixed molecular hypergraph

$$H = (V_H, E_H, \iota, \ell_V^H, \ell_E^H),$$

where $\iota : E_H \to \mathcal{P}(V_H) \setminus \{\varnothing\}$ is the incidence map (each hyperedge e is realized as the nonempty set $\iota(e) \subseteq V_H$ of bond–nodes incident to that atom).

Definition 13 (Indiscernibility and rough operators on V_H and E_H)**.** Let $\sim_V$ be an equivalence relation on V_H (vertex indiscernibility) and $\sim_E$ an equivalence relation on E_H (hyperedge indiscernibility). For $X \subseteq V_H$ define

$$\underline{X} := \{ v \in V_H : [v]_{\sim_V} \subseteq X \}, \qquad \overline{X} := \{ v \in V_H : [v]_{\sim_V} \cap X \neq \varnothing \},$$

and for $Y \subseteq E_H$ define

$$\underline{Y} := \{ e \in E_H : [e]_{\sim_E} \subseteq Y \}, \qquad \overline{Y} := \{ e \in E_H : [e]_{\sim_E} \cap Y \neq \varnothing \}.$$

Notation 2 (Vertex–induced admissible edges)**.** For $X \subseteq V_H$ set

$$E_H[X] := \{ e \in E_H : \iota(e) \subseteq X \},$$

the family of hyperedges fully supported on X.

Definition 14 (Rough Molecular HyperGraph (RMHG))**.** A *Rough Molecular HyperGraph* is a tuple

$$\mathcal{H}_{\mathrm{RMHG}} = \Big(H, \sim_V, \sim_E, \underline{(\cdot)}, \overline{(\cdot)}\Big),$$

where H is as above and $\sim_V, \sim_E$ are as in Definition 13. Given any *crisp target* subhypergraph $S = (X, Y)$ with $X \subseteq V_H$ and $Y \subseteq E_H$ satisfying the admissibility constraint

$$Y \subseteq E_H[X] \quad (\text{i.e. } \iota(e) \subseteq X \ \text{ for all } e \in Y),$$

the *lower* and *upper rough molecular subhypergraphs* of S are defined by

$$\underline{S} := (\underline{X}, \underline{Y} \cap E_H[\underline{X}]), \qquad \overline{S} := (\overline{X}, \overline{Y} \cap E_H[\overline{X}]).$$

Remark 3. As in standard rough sets, for all $X \subseteq V_H$ and $Y \subseteq E_H$,

$$\underline{X} \subseteq X \subseteq \overline{X}, \qquad \underline{Y} \subseteq Y \subseteq \overline{Y},$$

with monotonicity and idempotence: if $X \subseteq X'$ then $\underline{X} \subseteq \underline{X'}$ and $\overline{X} \subseteq \overline{X'}$, and $\underline{\underline{X}} = \underline{X}$, $\overline{\overline{X}} = \overline{X}$ (similarly for Y).

Example 11 (Benzene in solution as a Rough Molecular HyperGraph)**.** We present a real–life RMHG where solution NMR makes ring positions chemically equivalent, so individual carbons and C–H bonds cannot be distinguished at the hypergraph level.

3.2.5. Underlying molecular hypergraph

Label the ring carbons $C_1, \ldots, C_6$ cyclically (indices modulo 6) and the hydrogens $H_1, \ldots, H_6$ with H_i attached to C_i. The bond–node set and atom–hyperedges are

$$V_H = \{ b_{C_iC_{i+1}} : i = 1, \ldots, 6 \} \cup \{ b_{C_iH_i} : i = 1, \ldots, 6 \},$$

$$E_H = \{ e_{C_i} : i = 1, \ldots, 6 \} \cup \{ e_{H_i} : i = 1, \ldots, 6 \},$$

with incidence map $\iota : E_H \to \mathcal{P}(V_H) \setminus \{\varnothing\}$ given by

$$\iota(e_{C_i}) = \{ b_{C_{i-1}C_i}, b_{C_iC_{i+1}}, b_{C_iH_i} \}, \qquad \iota(e_{H_i}) = \{ b_{C_iH_i} \}.$$

3.2.6. Indiscernibility relations (solution-phase equivalence)

At ambient conditions benzene's ring positions are chemically equivalent (on the NMR timescale). Model this with two equivalence relations:

$$\underbrace{b_{C_iC_{i+1}} \sim_V b_{C_jC_{j+1}} \text{ for all } i,j,}_{\text{all aromatic C–C bonds equivalent}} \qquad \underbrace{b_{C_iH_i} \sim_V b_{C_jH_j} \text{ for all } i,j,}_{\text{all C–H bonds equivalent}}$$

$$\underbrace{e_{C_i} \sim_E e_{C_j} \text{ for all } i,j,}_{\text{all carbon atoms equivalent}} \qquad \underbrace{e_{H_i} \sim_E e_{H_j} \text{ for all } i,j.}_{\text{all hydrogens equivalent}}$$

These $\sim_V, \sim_E$ are used in the rough operators of Definition 13.

3.2.7. A crisp target subhypergraph and admissibility

Suppose we attempt to isolate the *specific* carbon site C_1 and its incident bonds:

$$X := \{ b_{C_1C_2}, b_{C_6C_1}, b_{C_1H_1} \} \subseteq V_H, \qquad Y := \{ e_{C_1} \} \subseteq E_H.$$

Since $\iota(e_{C_1}) \subseteq X$, we have $Y \subseteq E_H[X]$, hence $S = (X, Y)$ is an admissible crisp subhypergraph (Definition 14).

3.2.8. Rough approximations (explicit computation)

Vertices (bond–nodes). The $\sim_V$–class containing $b_{C_1C_2}$ is all six C–C bonds, which is *not* contained in X; similarly the class of $b_{C_1H_1}$ is all six C–H bonds, also not contained in X. Therefore

$$\underline{X} = \{v \in V_H : [v]_{\sim_V} \subseteq X\} = \varnothing, \qquad \overline{X} = \{v \in V_H : [v]_{\sim_V} \cap X \neq \varnothing\} = V_H,$$

(all C–C and all C–H bonds appear in the upper set).

Hyperedges (atoms). The $\sim_E$–class of e_{C_1} is $\{e_{C_1}, \ldots, e_{C_6}\}$, which is not a subset of Y; hence

$$\underline{Y} = \{e \in E_H : [e]_{\sim_E} \subseteq Y\} = \varnothing, \qquad \overline{Y} = \{e \in E_H : [e]_{\sim_E} \cap Y \neq \varnothing\} = \{e_{C_1}, \ldots, e_{C_6}\}.$$

Induced-edge filters. Using $E_H[U] = \{e \in E_H : \iota(e) \subseteq U\}$,

$$E_H[\underline{X}] = E_H[\varnothing] = \varnothing, \qquad E_H[\overline{X}] = E_H[V_H] = E_H.$$

Rough subhypergraphs (Definition 14). Thus

$$\underline{S} = (\underline{X}, \underline{Y} \cap E_H[\underline{X}]) = (\varnothing, \varnothing),$$

$$\overline{S} = (\overline{X}, \overline{Y} \cap E_H[\overline{X}]) = (V_H, \{e_{C_1}, \ldots, e_{C_6}\}).$$

Because all ring carbons (and their incident bonds) are experimentally indistinguishable in solution, no *specific* carbon site can be asserted with certainty: the lower RMHG is empty. However, any carbon site is *possible*, so the upper RMHG contains all six carbon hyperedges and, correspondingly, all C–C and C–H bonds. This matches the real benzene scenario where site labels are averaged by rapid symmetry–preserving dynamics on the observation timescale.

Lemma 7 (Well-defined rough subhypergraphs and sandwiching)**.** *For every admissible crisp $S = (X, Y)$ in H, the pairs $\underline{S}$ and $\overline{S}$ of Definition 14 are crisp subhypergraphs of H and satisfy*

$$\underline{S} \subseteq S \subseteq \overline{S} \quad \textit{(componentwise on vertices and hyperedges).}$$

Proof. We prove the lower case; the upper case is analogous. Since $\underline{X} \subseteq X$ by Remark 3, every edge in $E_H[\underline{X}]$ has its incidence contained in $\underline{X}$. Intersecting with $\underline{Y}$ can only remove edges, so $\underline{Y} \cap E_H[\underline{X}] \subseteq E_H[\underline{X}]$, hence $\underline{S}$ is a subhypergraph with vertex set $\underline{X}$ and hyperedge set $\underline{Y} \cap E_H[\underline{X}]$.

For the sandwich, vertices: $\underline{X} \subseteq X$ is clear. Hyperedges: $\underline{Y} \subseteq Y$ and $E_H[\underline{X}] \subseteq E_H[X]$, therefore

$$\underline{Y} \cap E_H[\underline{X}] \ \subseteq \ Y \cap E_H[X] \ = \ Y,$$

using the admissibility $Y \subseteq E_H[X]$. Thus $\underline{S} \subseteq S$. Similarly $S \subseteq \overline{S}$ follows from $X \subseteq \overline{X}$, $Y \subseteq \overline{Y}$ and $E_H[X] \subseteq E_H[\overline{X}]$. □

Lemma 8 (Monotonicity and idempotence at the subhypergraph level)**.** *If $S_1 = (X_1, Y_1)$ and $S_2 = (X_2, Y_2)$ are admissible crisp subhypergraphs with $S_1 \subseteq S_2$ (componentwise), then*

$$\underline{S_1} \ \subseteq \ \underline{S_2}, \qquad \overline{S_1} \ \subseteq \ \overline{S_2}.$$

Moreover, $\underline{(\underline{S})} = \underline{S}$ and $\overline{(\overline{S})} = \overline{S}$.

Proof. From $X_1 \subseteq X_2$ and $Y_1 \subseteq Y_2$, Remark 3 gives $\underline{X_1} \subseteq \underline{X_2}$ and $\underline{Y_1} \subseteq \underline{Y_2}$, and $E_H[\underline{X_1}] \subseteq E_H[\underline{X_2}]$; hence $\underline{Y_1} \cap E_H[\underline{X_1}] \subseteq \underline{Y_2} \cap E_H[\underline{X_2}]$. The upper case is analogous. Idempotence follows from the idempotence of $\underline{(\cdot)}$ and $\overline{(\cdot)}$ on V_H and E_H together with $E_H[\underline{\underline{X}}] = E_H[\underline{X}]$ and $E_H[\overline{\overline{X}}] = E_H[\overline{X}]$. □

Theorem 4 (RMHG generalizes Molecular HyperGraphs)**.** *Let $H = (V_H, E_H, \iota, \ell_V^H, \ell_E^H)$ be a (crisp) molecular hypergraph. Equip V_H and E_H with the* discrete *equivalence relations*

$$\sim_V = \text{ equality on } V_H, \qquad \sim_E = \text{ equality on } E_H.$$

Then for every admissible crisp subhypergraph $S = (X, Y)$ one has

$$\underline{S} \ = \ S \ = \ \overline{S}.$$

Consequently the assignment

$$H \ \longmapsto \ \Big(H, \ =_{V_H}, \ =_{E_H}, \ \underline{(\cdot)}, \ \overline{(\cdot)}\Big),$$

embeds the category of molecular hypergraphs into RMHGs.

Proof. With discrete relations, each class is a singleton; for $X \subseteq V_H$ and $Y \subseteq E_H$,

$$\underline{X} = X = \overline{X}, \qquad \underline{Y} = Y = \overline{Y}.$$

Hence $\underline{S} = (X, \ Y \cap E_H[X]) = (X, Y) = S$ by admissibility $Y \subseteq E_H[X]$, and similarly $\overline{S} = S$. □

Lemma 9 (Graphs as 2–uniform hypergraphs)**.** *Let $G = (V, E)$ be a (molecular) graph. Its 2–uniform hypergraph image is*

$$\Psi(G) \ := \ (V_H := V, \ E_H := E, \ \iota(e) = e \subseteq V).$$

If $\sim_V$ is an equivalence relation on V and $\sim_E$ on E, then for any $X \subseteq V$ and $Y \subseteq E$,

$$E_H[X] \ = \ \{ e \in E : \ e \subseteq X \} \ =: \ E[X],$$

and the rough operators on V_H, E_H coincide with the graph rough operators on V, E.

Proof. By construction $\iota(e) = e$ and $|e| = 2$, so $e \in E_H[X]$ iff both endpoints of e lie in X, i.e. $e \in E[X]$. Identifying V_H with V and E_H with E, the equivalence classes and hence the lower/upper operators are identical. □

Theorem 5 (RMHG generalizes Rough Molecular Graphs)**.** *Let $G = (V, E)$ be a molecular graph equipped with vertex/edge indiscernibility relations $\sim_V$ on V and $\sim_E$ on E. Consider the RMHG*

$$\mathcal{H}_{\mathrm{RMHG}} \ = \ \Big(\Psi(G), \ \sim_V, \ \sim_E, \ \underline{(\cdot)}, \ \overline{(\cdot)}\Big),$$

where $\Psi(G)$ is as in Lemma 9. For every admissible crisp subgraph $S = (X, Y)$ of G (i.e. $Y \subseteq E[X]$), the lower/upper rough subhypergraphs computed in $\mathcal{H}_{\mathrm{RMHG}}$ equal the lower/upper rough subgraphs computed in the Rough Molecular Graph model of G:

$$\underline{S}^{(\mathrm{RMHG})} = \underline{S}^{(\mathrm{RMG})}, \qquad \overline{S}^{(\mathrm{RMHG})} = \overline{S}^{(\mathrm{RMG})}.$$

Proof. By Lemma 9, $E_H[X] = E[X]$ and the rough operators on V_H, E_H coincide with those on V, E. Therefore

$$\underline{S}^{(\mathrm{RMHG})} = (\underline{X},\ \underline{Y} \cap E_H[\underline{X}]) = (\underline{X},\ \underline{Y} \cap E[\underline{X}]) = \underline{S}^{(\mathrm{RMG})},$$

and similarly for the upper approximation. □

3.2.9. Rough molecular SuperHyperGraph

A hierarchical molecular superhypergraph with lower and upper approximations, modeling uncertain atomic interactions and nested molecular substructures under rough set theory.

Notation 3 (Admissible superedges over a vertex set)**.** For $X \subseteq V$ set

$$E[X] := \{\, e \in E : e \subseteq X \,\}.$$

Note that if $H^{(n)}$ satisfies the hierarchy constraint, then every $e \in E[X]$ still satisfies $\mathrm{rk}(u) < \mathrm{rk}(e)$ for all $u \in e$ (because e and its members are unchanged).

Definition 15 (Rough Molecular SuperHyperGraph (RMSHG))**.** Fix $n \in \mathbb{N}$ and an MSHG

$$H^{(n)} = (V, E, \ell_V, \ell_E).$$

A *Rough Molecular SuperHyperGraph* is a tuple

$$\mathcal{H}^{(n)}_{\mathrm{RMSHG}} := \left(H^{(n)}, \sim_V, \sim_E, \underline{(\cdot)}, \overline{(\cdot)}\right),$$

where $\sim_V$ and $\sim_E$ are as in the Definition. Given any *crisp target* subsuperhypergraph $S = (X, Y)$ with $X \subseteq V$, $Y \subseteq E$ and the admissibility constraint $Y \subseteq E[X]$, its *lower* and *upper* rough subsuperhypergraphs are

$$\underline{S} := (\,\underline{X},\ \underline{Y} \cap E[\underline{X}]\,), \qquad \overline{S} := (\,\overline{X},\ \overline{Y} \cap E[\overline{X}]\,).$$

Remark 4 (Basic properties on V and E)**.** For all $X \subseteq V$, $Y \subseteq E$:

$$\underline{X} \subseteq X \subseteq \overline{X}, \qquad \underline{Y} \subseteq Y \subseteq \overline{Y},$$

and the operators are monotone and idempotent:

$$X \subseteq X' \Rightarrow \underline{X} \subseteq \underline{X'} \text{ and } \overline{X} \subseteq \overline{X'}, \quad \underline{\underline{X}} = \underline{X},\ \overline{\overline{X}} = \overline{X},$$

and similarly for Y.

Example 12 (Acetic acid–water hydrogen bonding as an RMSHG)**.** We exhibit a concrete real–life Rough Molecular SuperHyperGraph (RMSHG) built on a solvated acetic acid CH_3COOH with one water molecule. Hydrogen bonds form and break stochastically; we capture this uncertainty by rough (lower/upper) approximations.

3.2.10. MSHG backbone (depth $n = 3$)

Let B be the finite set of covalent bonds

$$B = \{b_{C_1C_2}, b_{C_1H_{1a}}, b_{C_1H_{1b}}, b_{C_1H_{1c}}, b_{C_2O_d}\ (\mathrm{C{=}O}), b_{C_2O_s}\ (\mathrm{C\text{-}O}), b_{O_sH_a}, b_{O_wH_{w1}}, b_{O_wH_{w2}}\},$$

where O_d is the carbonyl oxygen, O_s the hydroxyl oxygen, and (O_w, H_{w1}, H_{w2}) the water atoms.

Rank-1 *(atom) supervertices* are the sets of bonds incident to each atom:

$$v_{C_1} = \{b_{C_1C_2}, b_{C_1H_{1a}}, b_{C_1H_{1b}}, b_{C_1H_{1c}}\}, \quad v_{C_2} = \{b_{C_1C_2}, b_{C_2O_d}, b_{C_2O_s}\},$$
$$v_{O_d} = \{b_{C_2O_d}\}, \quad v_{O_s} = \{b_{C_2O_s}, b_{O_sH_a}\}, \quad v_{H_a} = \{b_{O_sH_a}\},$$
$$v_{O_w} = \{b_{O_wH_{w1}}, b_{O_wH_{w2}}\}, \quad v_{H_{w1}} = \{b_{O_wH_{w1}}\}, \quad v_{H_{w2}} = \{b_{O_wH_{w2}}\}.$$

Rank-2 *(group) supervertices* are sets of atoms:

$$A_{\mathrm{CH_3}} = \{v_{C_1}, v_{H_{1a}}, v_{H_{1b}}, v_{H_{1c}}\}, \quad A_{\mathrm{C=O}} = \{v_{C_2}, v_{O_d}\}, \quad A_{\mathrm{OH}} = \{v_{O_s}, v_{H_a}\}, \quad A_{\mathrm{H_2O}} = \{v_{O_w}, v_{H_{w1}}, v_{H_{w2}}\}.$$

Put all rank-1 and rank-2 supervertices into V.

Rank-3 *(interaction) superedges* encode supramolecular relations between groups:

$$e_{\mathrm{HB}(w1\to C=O)} = \{A_{\mathrm{H_2O}}, A_{\mathrm{C=O}}\}, \qquad e_{\mathrm{HB}(w2\to C=O)} = \{A_{\mathrm{H_2O}}, A_{\mathrm{C=O}}\}.$$

(They differ by which water hydrogen donates transiently to O_d.) Set

$$E \supseteq \{\, e_{\mathrm{HB}(w1\to C=O)},\, e_{\mathrm{HB}(w2\to C=O)} \,\} \subseteq \mathcal{P}^2(V),$$

so that $\mathrm{rk}(u) < \mathrm{rk}(e)$ holds for all incidences $u \in e$.

3.2.11. Indiscernibility (roughness) and occupancy data

Define equivalence relations capturing chemical indistinguishability:

$$v_{H_{w1}} \sim_V v_{H_{w2}}, \qquad v_{H_{1a}} \sim_V v_{H_{1b}} \sim_V v_{H_{1c}},$$

and, on interaction superedges,

$$e_{\mathrm{HB}(w1\to C=O)} \;\sim_E\; e_{\mathrm{HB}(w2\to C=O)} \qquad (\text{“water} \to \text{carbonyl” type}).$$

From molecular dynamics snapshots $\{\omega_j\}_{j=1}^N$, define empirical occupancies

$$p(e) := \frac{1}{N}\,\#\{j :\; e \in Y(\omega_j)\}.$$

Suppose

$$p\big(e_{\mathrm{HB}(w1\to C=O)}\big) = 0.82, \qquad p\big(e_{\mathrm{HB}(w2\to C=O)}\big) = 0.35.$$

Fix the certainty threshold $\tau = 0.5$ and let the *crisp target*

$$X := \{A_{\mathrm{C=O}}, A_{\mathrm{OH}}, A_{\mathrm{H_2O}}\} \subseteq V, \qquad Y_\tau := \{\, e \in E :\; p(e) \geq \tau \,\} = \{\, e_{\mathrm{HB}(w1\to C=O)} \,\}.$$

By construction $Y_\tau \subseteq E[X]$.

3.2.12. Rough approximations (explicit computation)

With $\sim_V$ as above, no nontrivial class lies entirely outside X, hence

$$\underline{X} = X = \overline{X}.$$

For superedges, the single $\sim_E$–class is

$$\big[e_{\mathrm{HB}(w1\to C=O)}\big]_{\sim_E} = \{\, e_{\mathrm{HB}(w1\to C=O)},\, e_{\mathrm{HB}(w2\to C=O)} \,\}.$$

Therefore, by the RMSHG Definition 15,

$$\underline{Y_\tau} = \{\, e \in E : \ [e]_{\sim_E} \subseteq Y_\tau \,\} = \varnothing \quad \text{since} \quad [e] \nsubseteq Y_\tau,$$

while

$$\overline{Y_\tau} = \{\, e \in E : \ [e]_{\sim_E} \cap Y_\tau \neq \varnothing \,\} = \{\, e_{\mathrm{HB}(w1\to C=O)},\ e_{\mathrm{HB}(w2\to C=O)} \,\}.$$

Intersecting with $E[\underline{X}] = E[X] = E[\overline{X}]$ yields the rough subsuperhypergraphs

$$\underline{S_\tau} = \big(\underline{X},\ \underline{Y_\tau} \cap E[\underline{X}]\big) = (X,\ \varnothing), \qquad \overline{S_\tau} = \big(\overline{X},\ \overline{Y_\tau} \cap E[\overline{X}]\big) = \big(X,\ \{e_{\mathrm{HB}(w1\to C=O)}, e_{\mathrm{HB}(w2\to C=O)}\}\big).$$

With $\tau = 0.5$, the *certain* (lower) RMSHG contains no water→carbonyl H–bond, because the entire indistinguishability class is not always present; the *possible* (upper) RMSHG contains both donor choices H_{w1} and H_{w2}, reflecting experimentally observed, rapidly exchanging hydrogen bonds in solution.

Example 13 (Formic acid dimer $(HCOOH)_2$ as a Rough Molecular SuperHyperGraph)**.** We model the well–known cyclic dimer of formic acid, stabilized by two *symmetric, rapidly exchanging* O–H$\cdots$O hydrogen bonds. The exchange renders donor/acceptor assignments indiscernible at coarse observational timescales (IR/NMR), which we capture via rough approximations on a Molecular SuperHyperGraph (MSHG).

3.2.13. MSHG backbone (depth $n = 3$)

Let B be the set of *covalent* bonds in two monomers A and B:

$$B = \{b_{C_A O_{Ad}}\ (\text{C=O}),\ b_{C_A O_{As}}\ (\text{C-O}),\ b_{O_{As} H_A}, b_{C_B O_{Bd}}\ (\text{C=O}),\ b_{C_B O_{Bs}}\ (\text{C-O}),\ b_{O_{Bs} H_B}\ \}.$$

Rank-1 *supervertices (atoms as sets of bonds).*

$$\begin{aligned} &v_{O_{Ad}} = \{b_{C_A O_{Ad}}\}, \quad v_{O_{As}} = \{b_{C_A O_{As}}, b_{O_{As} H_A}\}, \quad v_{H_A} = \{b_{O_{As} H_A}\}, \\ &v_{O_{Bd}} = \{b_{C_B O_{Bd}}\}, \quad v_{O_{Bs}} = \{b_{C_B O_{Bs}}, b_{O_{Bs} H_B}\}, \quad v_{H_B} = \{b_{O_{Bs} H_B}\}, \\ &\text{(carbon vertices omitted for brevity, not used below).} \end{aligned}$$

Rank-2 *supervertices (molecules as groups of atoms).*

$$A_1 = \{v_{O_{Ad}}, v_{O_{As}}, v_{H_A}\}, \qquad A_2 = \{v_{O_{Bd}}, v_{O_{Bs}}, v_{H_B}\}.$$

Let V contain all the above rank-1 and rank-2 supervertices.

Rank-3 *interaction superedges (hydrogen bonds).* To distinguish the two directions of the double H–bond, define

$$\begin{aligned} e_{\mathrm{HB}}^{A\to B} &= \{A_1, A_2,\ v_{H_A},\ v_{O_{Bd}}\}, \\ e_{\mathrm{HB}}^{B\to A} &= \{A_1, A_2,\ v_{H_B},\ v_{O_{Ad}}\}, \end{aligned}$$

so $e \in E \subseteq \mathcal{P}(V)$ has rank 3, while its members have rank 1 or 2, hence the hierarchy constraint $\mathrm{rk}(u) < \mathrm{rk}(e)$ holds.

3.2.14. Indiscernibility relations (chemical symmetry/exchange)

Fast donor/acceptor exchange and monomer symmetry are encoded by equivalence relations:

$$A_1 \sim_V A_2, \quad v_{H_A} \sim_V v_{H_B}, \quad v_{O_{Ad}} \sim_V v_{O_{Bd}}, \quad v_{O_{As}} \sim_V v_{O_{Bs}},$$

(and identity on other vertices); and on superedges

$$e_{\mathrm{HB}}^{A\to B} \ \sim_E\ e_{\mathrm{HB}}^{B\to A} \qquad \text{(“hydrogen bond between monomers” type).}$$

These $\sim_V, \sim_E$ are used in the rough operators of the Definition.

3.2.15. A crisp target subsuperhypergraph and admissibility

Suppose a probe aims to isolate the *specific* hydrogen bond from the A-monomer hydroxyl to the B-monomer carbonyl. Take

$$X = \{A_1, A_2, v_{H_A}, v_{O_{Bd}}\} \subseteq V, \qquad Y = \{\, e_{\mathrm{HB}}^{A\to B} \,\} \subseteq E.$$

Since $e_{\mathrm{HB}}^{A\to B} \subseteq X$, the admissibility $Y \subseteq E[X]$ (Notation 3) holds, so $S = (X, Y)$ is a valid crisp target.

3.2.16. Rough approximations (explicit computation)

Vertices. The $\sim_V$–classes intersecting X are

$$[A_1] = \{A_1, A_2\} \subseteq X, \quad [A_2] = \{A_1, A_2\} \subseteq X, \quad [v_{H_A}] = \{v_{H_A}, v_{H_B}\} \nsubseteq X, \quad [v_{O_{Bd}}] = \{v_{O_{Ad}}, v_{O_{Bd}}\} \nsubseteq X.$$

Hence

$$\underline{X} = \{A_1, A_2\}, \qquad \overline{X} = \{A_1, A_2, v_{H_A}, v_{H_B}, v_{O_{Ad}}, v_{O_{Bd}}\}.$$

Superedges. The class of $e_{\mathrm{HB}}^{A\to B}$ is $[e_{\mathrm{HB}}^{A\to B}]_{\sim_E} = \{e_{\mathrm{HB}}^{A\to B}, e_{\mathrm{HB}}^{B\to A}\}$, so

$$\underline{Y} = \varnothing, \qquad \overline{Y} = \{e_{\mathrm{HB}}^{A\to B}, e_{\mathrm{HB}}^{B\to A}\}.$$

Induced-edge filters. Using $E[U] = \{e \in E : \ e \subseteq U\}$,

$$E[\underline{X}] = E[\{A_1, A_2\}] = \varnothing \quad \text{(each HB also contains rank-1 members)},$$

$$E[\overline{X}] = E[\{A_1, A_2, v_{H_A}, v_{H_B}, v_{O_{Ad}}, v_{O_{Bd}}\}] \ \supseteq \ \{e_{\mathrm{HB}}^{A\to B}, e_{\mathrm{HB}}^{B\to A}\}.$$

Rough subsuperhypergraphs (Definition 15). Therefore

$$\underline{S} = (\underline{X}, \underline{Y} \cap E[\underline{X}]) = (\{A_1, A_2\}, \varnothing),$$

$$\overline{S} = (\overline{X}, \overline{Y} \cap E[\overline{X}]) = \Big(\{A_1, A_2, v_{H_A}, v_{H_B}, v_{O_{Ad}}, v_{O_{Bd}}\}, \ \{e_{\mathrm{HB}}^{A\to B}, e_{\mathrm{HB}}^{B\to A}\}\Big).$$

Because donor/acceptor roles exchange rapidly and the two monomers are indistinguishable, no *single* directed hydrogen bond can be asserted with certainty (the lower RMSHG contains none). However, both directed H–bond superedges are *possible* (the upper RMSHG contains both), faithfully representing the real cyclic dimer's symmetric, fluxional H–bond network.

Lemma 10 (Well-defined rough subsuperhypergraphs and sandwiching)**.** *For every admissible crisp $S = (X, Y)$ in $H^{(n)}$, the pairs $\underline{S}$ and $\overline{S}$ in Definition 15 are crisp subsuperhypergraphs of $H^{(n)}$ and*

$$\underline{S} \ \subseteq \ S \ \subseteq \ \overline{S} \quad \textit{(componentwise on vertices and superedges).}$$

Proof. We show the lower case; the upper case is analogous. By Remark 4, $\underline{X} \subseteq X$, hence every $e \in E[\underline{X}]$ has $e \subseteq \underline{X}$ and remains a valid superedge with the same rank relations. Intersecting with $\underline{Y}$ can only remove edges, so $\underline{Y} \cap E[\underline{X}] \subseteq E[\underline{X}]$, proving that $\underline{S}$ is a subsuperhypergraph on vertex set $\underline{X}$. For sandwiching, vertices: $\underline{X} \subseteq X$. Superedges: $\underline{Y} \subseteq Y$ and $E[\underline{X}] \subseteq E[X]$, thus

$$\underline{Y} \cap E[\underline{X}] \ \subseteq \ Y \cap E[X] \ = \ Y,$$

using admissibility $Y \subseteq E[X]$. Hence $\underline{S} \subseteq S$; the inclusion $S \subseteq \overline{S}$ follows similarly from $X \subseteq \overline{X}$, $Y \subseteq \overline{Y}$ and $E[X] \subseteq E[\overline{X}]$. $\square$

Lemma 11 (Monotonicity and idempotence at the substructure level)**.** *If $S_1 = (X_1, Y_1)$ and $S_2 = (X_2, Y_2)$ are admissible crisp subsuperhypergraphs with $S_1 \subseteq S_2$ (componentwise), then*

$$\underline{S_1} \ \subseteq \ \underline{S_2}, \qquad \overline{S_1} \ \subseteq \ \overline{S_2}.$$

Moreover, $\underline{(\underline{S})} = \underline{S}$ *and* $\overline{(\overline{S})} = \overline{S}$.

Proof. From $X_1 \subseteq X_2$ and $Y_1 \subseteq Y_2$, Remark 4 gives $\underline{X_1} \subseteq \underline{X_2}$ and $\underline{Y_1} \subseteq \underline{Y_2}$, and $E[\underline{X_1}] \subseteq E[\underline{X_2}]$; hence

$$\underline{Y_1} \cap E[\underline{X_1}] \subseteq \underline{Y_2} \cap E[\underline{X_2}].$$

The upper case is analogous. Idempotence follows from the idempotence of the rough operators on V and E together with $E[\underline{\underline{X}}] = E[\underline{X}]$ and $E[\overline{\overline{X}}] = E[\overline{X}]$. □

Theorem 6 (RMSHG generalizes Molecular SuperHyperGraphs). *Let* $H^{(n)} = (V, E, \ell_V, \ell_E)$ *be a (crisp) MSHG. Equip V and E with the* discrete *equivalence relations*

$$\sim_V = \text{ equality on } V, \qquad \sim_E = \text{ equality on } E.$$

Then for every admissible crisp subsuperhypergraph $S = (X, Y)$ *one has*

$$\underline{S} = S = \overline{S}.$$

Consequently,

$$H^{(n)} \longmapsto \left(H^{(n)}, =_V, =_E, \underline{(\cdot)}, \overline{(\cdot)}\right),$$

embeds the class of MSHGs into RMSHGs.

Proof. With discrete relations, every class is a singleton, so for all $X \subseteq V$, $Y \subseteq E$,

$$\underline{X} = X = \overline{X}, \qquad \underline{Y} = Y = \overline{Y}.$$

Therefore

$$\underline{S} = (\underline{X}, \underline{Y} \cap E[\underline{X}]) = (X, Y \cap E[X]) = (X, Y) = S,$$

by admissibility $Y \subseteq E[X]$. The same computation gives $\overline{S} = S$. □

Lemma 12 (Hypergraphs as depth-1 superhypergraphs, with roughness). *Let* $H = (V_H, E_H, \iota)$ *be a (bond–centric) molecular hypergraph with incidence* $\iota : E_H \to \mathcal{P}(V_H) \setminus \{\varnothing\}$. *Assume ι is injective (a standard situation in bond–centric models where bonds are unique). Define a depth-*1 *MSHG*

$$\Phi(H) := (V := V_H\ (\mathrm{rk} = 0),\ E := \{\iota(e) : e \in E_H\}\ (\mathrm{rk} = 1)).$$

Given equivalence relations $\sim_V$ *on* V_H *and* $\sim_E$ *on* E_H*, push them forward to V and E by*

$$v \sim_V^{\Phi} v' \iff v \sim_V v', \qquad A \sim_E^{\Phi} B \iff \exists\, e, f \in E_H : \iota(e) = A,\ \iota(f) = B,\ e \sim_E f.$$

If ι is injective, then $\sim_E^{\Phi}$ *is a well-defined equivalence on E. For every* $X \subseteq V_H$ *and* $Y \subseteq E_H$ *one has*

$$E[\,X\,] = \{\,\iota(e) : e \in E_H,\ \iota(e) \subseteq X\,\} = \iota\,(E_H[X]),$$

and the rough operators on (V, E) *coincide with those induced from* (V_H, E_H) *via* Φ.

Proof. If ι is injective, each $A \in E$ has a unique preimage $e \in E_H$, so $\sim_E^{\Phi}$ is well-defined and inherits reflexivity, symmetry, and transitivity from $\sim_E$. The identity $E[X] = \iota(E_H[X])$ holds by definition of E and $E_H[X]$. The formulas for lower/upper approximations transport through ι componentwise. □

Theorem 7 (RMSHG generalizes Rough Molecular HyperGraphs). *Let* $\mathcal{H}_{\mathrm{RMHG}} = \left(H, \sim_V, \sim_E, \underline{(\cdot)}, \overline{(\cdot)}\right)$ *be a Rough Molecular HyperGraph in the bond–centric presentation, and assume its incidence map ι is injective. Form the RMSHG*

$$\mathcal{H}_{\mathrm{RMSHG}}^{(1)} := \left(\Phi(H), \sim_V^{\Phi}, \sim_E^{\Phi}, \underline{(\cdot)}, \overline{(\cdot)}\right),$$

as in Lemma 12. Then for every admissible crisp subhypergraph $S = (X, Y)$ of H (i.e. $Y \subseteq E_H[X]$), the rough lower/upper subsuperhypergraphs computed in $\mathcal{H}^{(1)}_{\mathrm{RMSHG}}$ coincide with the rough lower/upper subhypergraphs of $\mathcal{H}_{\mathrm{RMHG}}$ under the identification by ι:

$$\underline{S}^{(\mathrm{RMSHG})} = \iota\left(\underline{S}^{(\mathrm{RMHG})}\right), \qquad \overline{S}^{(\mathrm{RMSHG})} = \iota\left(\overline{S}^{(\mathrm{RMHG})}\right).$$

Proof. By Lemma 12, $E[\underline{X}] = \iota(E_H[\underline{X}])$ and the pushed-forward equivalence relations yield the same lower/upper operators on V and E as on V_H and E_H after applying ι. Hence

$$\underline{S}^{(\mathrm{RMSHG})} = (\underline{X},\ \underline{Y} \cap E[\underline{X}]) = (\underline{X},\ \iota(\underline{Y} \cap E_H[\underline{X}])) = \iota\left(\underline{S}^{(\mathrm{RMHG})}\right),$$

and similarly for the upper approximation. □

3.3. Molecular graph neural networks

3.3.1. Molecular hypergraph neural networks

Molecular Graph Neural Networks are deep learning frameworks that operate on graphs, where atoms are treated as vertices and chemical bonds as edges, enabling the learning of molecular properties through iterative message passing [29,30]. Molecular Hypergraph Neural Networks extend this idea to hypergraphs, where hyperedges connect multiple atoms simultaneously, allowing the model to capture higher-order chemical interactions and more complex structural motifs than ordinary graphs [32,33,61].

Definition 16 (Molecular Hypergraph Neural Network (MHNN))**.** [32] Let a molecular hypergraph be $G = (V, E, H, L)$ with $n = |V|$ nodes (atoms), $m = |E|$ hyperedges (multi-atom relations), node features $H \in \mathbb{R}^{n \times d}$ and hyperedge features $L \in \mathbb{R}^{m \times d'}$. We write $v \in V$, $e \in E$, and $v \in e$ iff node v belongs to hyperedge e. A standard MHNN operates on the incidence bipartite graph between V and E and performs message passing with four differentiable maps f_1, f_2, f_3, f_4 (e.g. MLPs), one pair for $V \to E$ and one for $E \to V$.

For $t = 1, \ldots, T$, with hidden states $h_v^{(t)} \in \mathbb{R}^{d_h}$ for $v \in V$ and $\ell_e^{(t)} \in \mathbb{R}^{d_\ell}$ for $e \in E$, define the layer updates:

$$m^{(t)}_{v \to e} = \sum_{v \in e} f_1\big(h_v^{(t-1)},\ \ell_e^{(t-1)}\big), \qquad \ell_e^{(t)} = f_2\big(\ell_e^{(t-1)},\ m^{(t)}_{v \to e}\big), \tag{1}$$

$$m^{(t)}_{e \to v} = \sum_{e:\, v \in e} f_3\big(\ell_e^{(t)},\ h_v^{(t-1)}\big), \qquad h_v^{(t)} = f_4\big(h_v^{(t-1)},\ m^{(t)}_{e \to v}\big). \tag{2}$$

The initial states are obtained from the input features, e.g. $h_v^{(0)} = \phi_V(H_v)$ and $\ell_e^{(0)} = \phi_E(L_e)$ for suitable encoders ϕ_V, ϕ_E. After T layers, a permutation-invariant readout produces the graph-level prediction

$$\hat{y} = \mathrm{MLP}\Big(\sum_{v \in V} h_v^{(T)},\ \sum_{e \in E} \ell_e^{(T)}\Big), \tag{3}$$

optionally restricting the second sum to higher-order hyperedges.

An MHNN instance is the tuple

$$\mathrm{MHNN}(G; \theta) = \big(f_1, f_2, f_3, f_4, \phi_V, \phi_E, \mathrm{MLP}, T\big)_\theta,$$

trained by minimizing a task loss $\mathcal{L}(\hat{y}, y)$ (e.g. MSE or cross-entropy) over data. The formulation is order-agnostic in $|e|$ and supports variable-order hyperedges via the bipartite message passing (1)–(2).

Example 14 (Chemistry: An MHNN for Ethanol CH_3CH_2OH Solubility)**.** We instantiate Definition MHNN on a small molecular hypergraph whose nodes are atoms and whose hyperedges are functional groups (multi–atom relations).

Molecular hypergraph.

Let the node (atom) set be

$$V = \{C_1,\ H_{1a}, H_{1b}, H_{1c},\ C_2,\ H_{2a}, H_{2b},\ O,\ H_O\}.$$

Define three hyperedges (functional groups)

$$\begin{aligned} e_{\mathrm{CH_3}} &= \{C_1, H_{1a}, H_{1b}, H_{1c}\}, \\ e_{\mathrm{CH_2}} &= \{C_2, H_{2a}, H_{2b}\}, \\ e_{\mathrm{OH}} &= \{O, H_O, C_2\}, \end{aligned}$$

so $E = \{e_{\mathrm{CH_3}}, e_{\mathrm{CH_2}}, e_{\mathrm{OH}}\} \subset \mathcal{P}(V) \setminus \{\varnothing\}$ and incidence is given by $v \in e$ as listed. This hypergraph captures higher–order (group–level) interactions beyond pairwise bonds.

3.3.2. Features

Let $H \in \mathbb{R}^{|V|\times d}$ encode atom descriptors (e.g. one–hot element, degree, hybridization) and $L \in \mathbb{R}^{|E|\times d'}$ encode group descriptors (e.g. group type, polarity proxies). We initialize hidden states by encoders $h_v^{(0)} = \phi_V(H_v) \in \mathbb{R}^{d_h}$ and $\ell_e^{(0)} = \phi_E(L_e) \in \mathbb{R}^{d_\ell}$.

3.3.3. One MHNN layer

For $t = 1$, messages from atoms to hyperedges are

$$m_{v\to e}^{(1)} \;=\; \sum_{v\in e} f_1\big(h_v^{(0)}, \ell_e^{(0)}\big),$$

so concretely

$$\begin{aligned} m_{v\to e_{\mathrm{CH_3}}}^{(1)} &= f_1(h_{\mathrm{C_1}}^{(0)}, \ell_{e_{\mathrm{CH_3}}}^{(0)}) + f_1(h_{H_{1a}}^{(0)}, \ell_{e_{\mathrm{CH_3}}}^{(0)}) + f_1(h_{H_{1b}}^{(0)}, \ell_{e_{\mathrm{CH_3}}}^{(0)}) + f_1(h_{\mathrm{H_{1c}}}^{(0)}, \ell_{e_{\mathrm{CH_3}}}^{(0)}), \\ m_{v\to e_{\mathrm{CH_2}}}^{(1)} &= f_1(h_{\mathrm{C_2}}^{(0)}, \ell_{e_{\mathrm{CH_2}}}^{(0)}) + f_1(h_{H_{2a}}^{(0)}, \ell_{e_{\mathrm{CH_2}}}^{(0)}) + f_1(h_{H_{2b}}^{(0)}, \ell_{e_{\mathrm{CH_2}}}^{(0)}), \\ m_{v\to e_{\mathrm{OH}}}^{(1)} &= f_1(h_{\mathrm{O}}^{(0)}, \ell_{e_{\mathrm{OH}}}^{(0)}) + f_1(h_{H_O}^{(0)}, \ell_{e_{\mathrm{OH}}}^{(0)}) + f_1(h_{\mathrm{C_2}}^{(0)}, \ell_{e_{\mathrm{OH}}}^{(0)}). \end{aligned}$$

Hyperedge updates aggregate these messages:

$$\ell_e^{(1)} = f_2\big(\ell_e^{(0)},\, m_{v\to e}^{(1)}\big) \qquad (e \in E).$$

Messages back to atoms sum over incident hyperedges:

$$m_{e\to v}^{(1)} \;=\; \sum_{e:\, v\in e} f_3\big(\ell_e^{(1)}, h_v^{(0)}\big),$$

and atoms update via

$$h_v^{(1)} \;=\; f_4\big(h_v^{(0)},\, m_{e\to v}^{(1)}\big) \qquad (v \in V).$$

For example, atom C_2 lies in two hyperedges ($e_{\mathrm{CH_2}}$ and e_{OH}), hence

$$m_{e\to \mathrm{C_2}}^{(1)} = f_3\big(\ell_{e_{\mathrm{CH_2}}}^{(1)}, h_{\mathrm{C_2}}^{(0)}\big) + f_3\big(\ell_{e_{\mathrm{OH}}}^{(1)}, h_{\mathrm{C_2}}^{(0)}\big), \qquad h_{\mathrm{C_2}}^{(1)} = f_4\big(h_{\mathrm{C_2}}^{(0)}, m_{e\to \mathrm{C_2}}^{(1)}\big).$$

3.3.4. Prediction and task

After T layers, a permutation–invariant readout produces a graph–level representation for a property such as aqueous solubility (logS):

$$\hat{y} \;=\; \mathrm{MLP}\Big(\sum_{v\in V} h_v^{(T)},\ \sum_{e\in E} \ell_e^{(T)} \Big).$$

The MHNN instance

$$\mathrm{MHNN}(G;\theta) = \big(f_1, f_2, f_3, f_4, \phi_V, \phi_E, \mathrm{MLP}, T\big)_\theta,$$

is trained by minimizing $\mathcal{L}(\hat{y}, y)$ over molecules with experimental logS labels. Function f_3 and f_4 let polar group signals from e_{OH} inform C_2 and thereby propagate hydrogen–bonding capacity to the global readout, while e_{CH_3} contributes hydrophobic context, enabling the network to learn the balance that governs solubility.

3.3.5. Molecular SuperHyperGraph neural networks

We design a message–passing architecture that operates on a Molecular n–SuperHyperGraph (MSHG) in the sense of Definition 6. Messages alternate between (super)vertices and (super)edges and respect the hierarchy constraint $\mathrm{rk}(u) < \mathrm{rk}(e)$ for every incidence $u \in e$.

Definition 17 (Molecular SuperHyperGraph Neural Network (MSHNN))**.** Let $H^{(n)} = (V, E, \ell_V, \ell_E)$ be an MSHG with rank map $\mathrm{rk} : V \cup E \to \{0, 1, \ldots, n\}$, where $\mathrm{rk}(u) < \mathrm{rk}(e)$ whenever $u \in e \in E$. Fix input features $x_v \in \mathbb{R}^{d_V}$ for $v \in V$ and $x_e \in \mathbb{R}^{d_E}$ for $e \in E$ (derived from labels ℓ_V, ℓ_E). An *MSHNN* consists of:

- encoders $\phi_V : \mathbb{R}^{d_V} \to \mathbb{R}^h$ and $\phi_E : \mathbb{R}^{d_E} \to \mathbb{R}^h$,
- (shared) message maps $f_{VE}, f_{EV} : \mathbb{R}^h \times \mathbb{R}^h \times \mathbb{R}^r \to \mathbb{R}^h$, and update maps $U_V, U_E : \mathbb{R}^h \times \mathbb{R}^h \to \mathbb{R}^h$,
- a rank embedding $\rho : \{0, 1, \ldots, n\} \to \mathbb{R}^r$,
- a permutation–invariant readout $\mathrm{RO} : \big(\mathbb{R}^h\big)^{|V|} \times \big(\mathbb{R}^h\big)^{|E|} \to \mathbb{R}^o$.

Initialize hidden states $h_v^{(0)} = \phi_V(x_v)$ and $s_e^{(0)} = \phi_E(x_e)$. For layers $t = 1, \ldots, T$, perform the alternating updates

$$\text{(V}\to\text{E messages)} \qquad M_{V\to e}^{(t)} = \sum_{u \in e} f_{VE}\big(h_u^{(t-1)},\, s_e^{(t-1)},\, \rho(\mathrm{rk}(u)) \oplus \rho(\mathrm{rk}(e))\big), \tag{4}$$

$$\text{(E update)} \qquad s_e^{(t)} = U_E\big(s_e^{(t-1)},\, M_{V\to e}^{(t)}\big), \tag{5}$$

$$\text{(E}\to\text{V messages)} \qquad M_{E\to v}^{(t)} = \sum_{e:\, v \in e} f_{EV}\big(s_e^{(t)},\, h_v^{(t-1)},\, \rho(\mathrm{rk}(e)) \oplus \rho(\mathrm{rk}(v))\big), \tag{6}$$

$$\text{(V update)} \qquad h_v^{(t)} = U_V\big(h_v^{(t-1)},\, M_{E\to v}^{(t)}\big), \tag{7}$$

where $\oplus$ denotes concatenation and the sums are over the (super)incidence relation of $H^{(n)}$. After T layers, produce the graph–level prediction

$$\hat{y} \;=\; \mathrm{RO}\Big(\{h_v^{(T)} : v \in V\},\; \{s_e^{(T)} : e \in E\}\Big),$$

and train parameters θ (of $\phi_V, \phi_E, f_{VE}, f_{EV}, U_V, U_E, \rho, \mathrm{RO}$) by minimizing a task loss $\mathcal{L}(\hat{y}, y)$.

Theorem 8 (MSHNN reduces to MHNN at depth n=1)**.** *Let $H^{(1)} = (V, E, \ell_V, \ell_E)$ be an MSHG of depth 1. Consider the Molecular HyperGraph $H_{\mathrm{HG}} = (V, E, \iota)$ obtained by viewing each $e \in E$ as a hyperedge on its member set (so $\iota(e) = e \subseteq V$). Fix an MSHNN on $H^{(1)}$ with the following choices:*

$$\rho(0) = \mathbf{0},\ \rho(1) = \mathbf{0},$$

$$f_{VE}(a, b, \cdot) = f_1(a, b), \quad U_E(b, c) = f_2(b, c),$$

$$f_{EV}(a, b, \cdot) = f_3(a, b), \quad U_V(b, c) = f_4(b, c),$$

for some differentiable f_1, f_2, f_3, f_4 (e.g. MLPs), and take the readout $\mathrm{RO}(H^{(T)}, S^{(T)}) = \mathrm{MLP}\big(\sum_v h_v^{(T)},\, \sum_e s_e^{(T)}\big)$. Then the layer updates (4)–(7) become exactly the standard Molecular HyperGraph Neural Network (MHNN) updates on H_{HG}:

$$\underbrace{\sum_{u\in e} f_1\big(h_u^{(t-1)}, s_e^{(t-1)}\big)}_{=:m_{v\to e}^{(t)}} \Longrightarrow s_e^{(t)} = f_2\big(s_e^{(t-1)}, m_{v\to e}^{(t)}\big),$$

$$\underbrace{\sum_{e:\,v\in e} f_3\big(s_e^{(t)}, h_v^{(t-1)}\big)}_{=:m_{e\to v}^{(t)}} \Longrightarrow h_v^{(t)} = f_4\big(h_v^{(t-1)}, m_{e\to v}^{(t)}\big),$$

with identical initialization and readout. Hence MSHNNs strictly generalize MHNNs.

Proof. At depth 1, every member u of a superedge e has $\mathrm{rk}(u) = 0$ and $\mathrm{rk}(e) = 1$. With $\rho(0) = \rho(1) = \mathbf{0}$, the rank–embedding argument passed to f_{VE}, f_{EV} is constant and can be absorbed into the learned bias terms, reducing (4)–(7) to

$$M_{V\to e}^{(t)} = \sum_{u\in e} f_1(h_u^{(t-1)}, s_e^{(t-1)}), \quad s_e^{(t)} = f_2(s_e^{(t-1)}, M_{V\to e}^{(t)}),$$

$$M_{E\to v}^{(t)} = \sum_{e:\,v\in e} f_3(s_e^{(t)}, h_v^{(t-1)}), \quad h_v^{(t)} = f_4(h_v^{(t-1)}, M_{E\to v}^{(t)}).$$

These are exactly the MHNN equations on the hypergraph H_{HG} (the incidence relation is the same). Initialization and readout coincide by construction, so all layerwise states and the final prediction match for all $t = 0, \dots, T$. Therefore MSHNN reduces to MHNN when n=1 under the stated parameter tying. □

3.4. Multipolar molecular models

In this subsection, we describe Multipolar Molecular Models.

3.4.1. Multipolar molecular graph

Multipolar Molecular Graphs are molecular graphs that incorporate spherical-harmonic multipole coefficients, allowing atomic positions and weights to be represented in a mathematically rigorous way that yields rotationally invariant molecular shape descriptors.

Definition 18 (Molecular graph with geometry and weights)**.** A *molecular graph with geometry* is a tuple

$$G = (V, E, \lambda_V, \lambda_E, \mathbf{r}, w),$$

where V is a finite set of atoms, $E \subseteq \{\{u, v\} : u \neq v \in V\}$ is the bond set, $\lambda_V : V \to \Sigma_V$ and $\lambda_E : E \to \Sigma_E$ are (optional) atom/bond labels, $\mathbf{r} : V \to \mathbb{R}^3$ assigns a 3D position to each atom, and $w : V \to \mathbb{R}_{>0}$ assigns a positive weight (e.g. mass, charge, or a residue property score).

Definition 19 (Centered coordinates)**.** Let $W := \sum_{v\in V} w(v)$ and the weighted centroid $\mathbf{r}_c := W^{-1} \sum_{v\in V} w(v)\,\mathbf{r}(v)$. For each $v \in V$ set $\widehat{\mathbf{r}}(v) := \mathbf{r}(v) - \mathbf{r}_c$ and write $\widehat{\mathbf{r}}(v)$ in spherical coordinates $r_v := \|\widehat{\mathbf{r}}(v)\|$, $(\theta_v, \phi_v) \in [0, \pi] \times [0, 2\pi)$.

Definition 20 (Multipole tensors of a molecular graph)**.** For each integer rank $l \geq 0$ and order $m \in \{-l, \dots, l\}$ define the *(complex) multipole coefficients* of G by

$$Q_{lm}(G) \;:=\; \sum_{v\in V} w(v)\, r_v^l\, Y_{lm}(\theta_v, \phi_v)^*,$$

where Y_{lm} are the (Condon–Shortley) spherical harmonics and * is complex conjugation. Let $Q_l(G) := (Q_{l,-l}, \dots, Q_{l,l}) \in \mathbb{C}^{2l+1}$ and $P_l(G) := \sum_{m=-l}^{l} |Q_{lm}(G)|^2$ (the *power spectrum* at rank l).

Definition 21 (Multipolar Molecular Graph (MMG))**.** Fix a truncation $L \in \mathbb{N}$. A *Multipolar Molecular Graph of order* L is the tuple

$$\mathsf{MMG}_L(G) \;=\; \Big(G,\ \{Q_l(G)\}_{l=0}^{L}\Big),$$

computed from a molecular graph with geometry and weights G as above. When a rotation–invariant summary is desired we use $\{P_l(G)\}_{l=0}^{L}$ instead of the coefficient vectors.

Example 15 (Acetone $(CH_3)_2CO$: strong carbonyl dipole)**.** *Molecular graph and geometry.* Let $G = (V, E, \mathbf{r}, w)$ be the atom–bond graph of acetone with

$$V = \{C_{\text{carbonyl}}, O, C_{m1}, C_{m2}, \text{ H's on each methyl}\}, \quad E = \{\{C_{\text{carbonyl}}, O\}, \{C_{\text{carbonyl}}, C_{m1}\}, \{C_{\text{carbonyl}}, C_{m2}\}, \text{ C–H edges}\}.$$

Use 3D coordinates $\mathbf{r}(v)$ from experiment or a gas–phase optimization (the carbonyl lies approximately in a plane; the molecular dipole aligns along the C=O axis).

Weights and centering. Choose chemically meaningful positive weights, e.g. $w(v) = |q(v)|$ from a consistent partial–charge model (RESP or AM1–BCC). Define

$$\mathbf{r}_c = \left(\sum_v w(v)\right)^{-1} \sum_v w(v)\,\mathbf{r}(v), \qquad \hat{\mathbf{r}}(v) = \mathbf{r}(v) - \mathbf{r}_c,$$

and the multipoles $Q_{lm}(G)$, $P_l(G)$.

Directional signature (dominant dipole). Because electron density and charges are polarized along C=O, the weighted configuration has a nonzero *dipole* order:

$$Q_{1m}(G) = \sum_{v \in V} w(v)\, r_v\, Y_{1m}(\theta_v, \phi_v)^* \;\neq\; 0,$$

with the vectorial content aligned to the carbonyl axis (up to rotation of the entire molecule). Higher orders ($l \geq 2$), especially P_2, encode the flattening of the environment around the carbonyl and the approximate C_2 symmetry imposed by the two methyl groups. The collection $\{P_l(G)\}_{l \leq L}$ thus provides a rotation–invariant fingerprint distinguishing strongly polar carbonyls nonpolar rings like benzene, while remaining faithful to the underlying graph G. Figure 7 illustrates the molecular graph of acetone together with the weighted centroid $\mathbf{r}_c$ and the dominant dipole direction along the C=O axis.

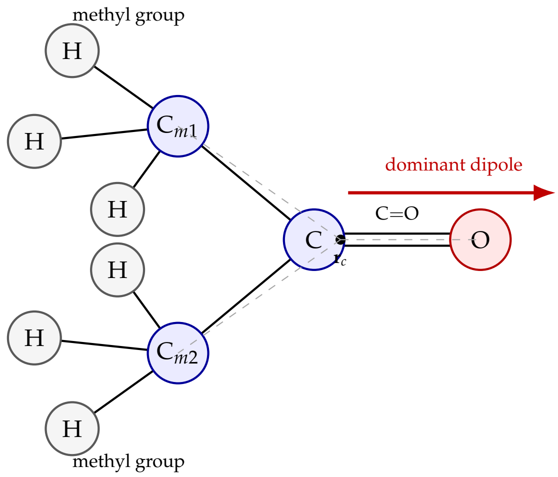


Figure 7. A schematic illustration of acetone $(CH_3)_2CO$ as a multipolar molecular graph. The weighted centroid $\mathbf{r}_c$ is shown together with the dominant dipole direction aligned with the carbonyl axis

Theorem 9 (Basic invariances of the MMG)**.** *Let G be a molecular graph with geometry and weights, and let $R \in SO(3)$ and $\mathbf{t} \in \mathbb{R}^3$. Denote by $G^{(R,\mathbf{t})}$ the graph obtained by replacing $\mathbf{r}(v)$ with $R\mathbf{r}(v) + \mathbf{t}$ for all $v \in V$, leaving $V, E, \lambda_V, \lambda_E, w$ unchanged. Then:*

(i) *(Translation invariance after centering)* $Q_{lm}\left(G^{(I,\mathbf{t})}\right) = Q_{lm}(G)$ *for all* l, m.

(ii) *(Rotation covariance) For each* l *there exists a unitary Wigner matrix* $D^{(l)}(R) \in \mathbb{C}^{(2l+1)\times(2l+1)}$ *such that*

$$Q_l\left(G^{(R,\mathbf{0})}\right) \;=\; D^{(l)}(R)\, Q_l(G).$$

In particular, the power spectrum *is rotation invariant:* $P_l\left(G^{(R,\mathbf{0})}\right) = P_l(G)$ *for all* l.

Hence the family $\{P_l(G)\}_{l\le L}$ *(and any function of it) is invariant under rigid motions, and* MMG_L *represents molecular shape/ property distributions up to rotation and translation.*

Proof. (i) Centering subtracts the weighted centroid; adding a common translation $\mathbf{t}$ does not change the centered vectors $\widehat{\mathbf{r}}(v) = \mathbf{r}(v) - \mathbf{r}_c$, so each Q_{lm} is unchanged.

(ii) Under a rotation R, the angles (θ_v, ϕ_v) of each centered vector transform by R, and the spherical harmonics obey $Y_{lm}(R\cdot\Omega) = \sum_{m'=-l}^{l} D^{(l)}_{mm'}(R)\, Y_{lm'}(\Omega)$, with $D^{(l)}$ unitary. Substituting into the definition of Q_{lm} yields $Q_l \mapsto D^{(l)}(R)Q_l$. Unitarity implies $P_l = \|Q_l\|_2^2$ is preserved. □

3.4.2. Multipolar molecular HyperGraph

Multipolar Molecular HyperGraphs extend this idea to hypergraphs, in which atom-hyperedges and bond-nodes carry both weights and multipole expansions, thereby capturing multi-atom interactions and structural motifs with built-in geometric invariance.

Definition 22 (Molecular hypergraph with geometry and weights)**.** A *molecular hypergraph with geometry* is a tuple

$$H = (V_H, E_H, \iota, \lambda_V, \lambda_E, \mathbf{r}_V, \mathbf{r}_E, w_V, w_E),$$

where V_H is a finite set of vertices (e.g. atoms or bonds), E_H a finite set of hyperedges, $\iota : E_H \to \mathcal{P}(V_H) \setminus \{\varnothing\}$ is the incidence map, λ_V, λ_E are optional labels, $\mathbf{r}_V : V_H \to \mathbb{R}^3$, $\mathbf{r}_E : E_H \to \mathbb{R}^3$ assign 3D positions (e.g. atomic coordinates for vertices, geometric representatives for hyperedges), and $w_V : V_H \to \mathbb{R}_{>0}$, $w_E : E_H \to \mathbb{R}_{>0}$ assign positive weights (e.g. masses/charges/importance).

Definition 23 (Carrier, centering, and multipoles)**.** Let the *carrier* $\mathsf{Car} \subseteq V_H \cup E_H$ be one of V_H, E_H, or $V_H \cup E_H$. Define weights and positions on Car by

$$w(a) := \begin{cases} w_V(a), & a \in V_H, \\ w_E(a), & a \in E_H, \end{cases} \qquad \mathbf{r}(a) := \begin{cases} \mathbf{r}_V(a), & a \in V_H, \\ \mathbf{r}_E(a), & a \in E_H. \end{cases}$$

Let $W := \sum_{a\in\mathsf{Car}} w(a)$ and the weighted centroid $\mathbf{r}_c := W^{-1} \sum_{a\in\mathsf{Car}} w(a)\,\mathbf{r}(a)$. Set $\widehat{\mathbf{r}}(a) := \mathbf{r}(a) - \mathbf{r}_c$ and write $r_a := \|\widehat{\mathbf{r}}(a)\|$ and (θ_a, ϕ_a) for its spherical angles. For each rank $l \ge 0$ and order $m \in \{-l, \dots, l\}$ define the (complex) multipoles

$$Q_{lm}(H, \mathsf{Car}) \;:=\; \sum_{a\in\mathsf{Car}} w(a)\, r_a^l\, Y_{lm}(\theta_a, \phi_a)^*,$$

$Q_l(H, \mathsf{Car}) = (Q_{l,-l}, \dots, Q_{l,l})$, and the power spectrum $P_l(H, \mathsf{Car}) = \sum_{m=-l}^{l} |Q_{lm}(H, \mathsf{Car})|^2$.

Definition 24 (Multipolar molecular HyperGraph (MpMHG))**.** Fix an order $L \in \mathbb{N}$ and a carrier choice $\mathsf{Car} \in \{V_H,\ E_H,\ V_H \cup E_H\}$. The *Multipolar Molecular HyperGraph of order* L is

$$\mathsf{MpMHG}_L(H, \mathsf{Car}) \;:=\; \left(H,\ \{Q_l(H, \mathsf{Car})\}_{l=0}^{L}\right).$$

When rotation invariance is preferred, we use $\{P_l(H, \mathsf{Car})\}_{l=0}^{L}$.

Example 16 (Aspirin (acetylsalicylic acid) with functional-group hyperedges)**.** *Atom-centric hypergraph.* Let $H = (V_H, E_H, \iota)$ be the *atom-centric* molecular hypergraph of aspirin. Enumerate atoms by

$$V_H = \{ \mathrm{C}_1, \ldots, \mathrm{C}_6 \text{ (aromatic)}, \mathrm{C}_{\mathrm{carb}}, \mathrm{C}_{\mathrm{acet}}, \mathrm{O}_{\mathrm{acid}}, \mathrm{O}_{\mathrm{ester1}}, \mathrm{O}_{\mathrm{ester2}}, \mathrm{H}_1, \ldots, \mathrm{H}_m \},$$

with Cartesian coordinates $\mathbf{r}_V : V_H \to \mathbb{R}^3$ taken from an X-ray or DFT structure. Define functional-group hyperedges

$$\begin{aligned} e_{\mathrm{ring}} &= \{\mathrm{C}_1, \ldots, \mathrm{C}_6, \text{ring-attached H's}\}, \\ e_{\mathrm{carboxyl}} &= \{\mathrm{C}_{\mathrm{carb}}, \mathrm{O}_{\mathrm{acid}}, \mathrm{O}_{\mathrm{ester1}}\}, \\ e_{\mathrm{ester}} &= \{\mathrm{C}_{\mathrm{acet}}, \mathrm{O}_{\mathrm{ester1}}, \mathrm{O}_{\mathrm{ester2}}\}, \\ e_{\mathrm{methyl}} &= \{\text{acetyl methyl C and its H's}\}, \end{aligned} \qquad E_H = \{e_{\mathrm{ring}}, e_{\mathrm{carboxyl}}, e_{\mathrm{ester}}, e_{\mathrm{methyl}}\},$$

with incidence $\iota(e) = e \subseteq V_H$.

Geometry and weights. Let atomic weights $w_V(v) := |q(v)|$ be the absolute partial charges (e.g. RESP). Define hyperedge representatives and weights by barycenters and sums:

$$\mathbf{r}_E(e) := \frac{1}{|e|} \sum_{v \in e} \mathbf{r}_V(v), \qquad w_E(e) := \sum_{v \in e} w_V(v).$$

Multipoles and carrier. Choose carrier $\mathsf{Car} = E_H$ (group-level description) and order $L \in \mathbb{N}$. Let $W = \sum_{a \in \mathsf{Car}} w(a)$, $\mathbf{r}_c = W^{-1} \sum_a w(a)\mathbf{r}(a)$, $\widehat{\mathbf{r}}(a) = \mathbf{r}(a) - \mathbf{r}_c$, and define for $l \geq 0$, $m = -l, \ldots, l$:

$$Q_{lm}\,(H, \mathsf{Car} = E_H) = \sum_{e \in E_H} w_E(e)\, \|\widehat{\mathbf{r}}(e)\|^l\, Y_{lm}(\theta_e, \phi_e)^*.$$

The associated MpMHG is

$$\mathsf{MpMHG}_L\,(H, E_H) = \left(H, \{Q_l(H, E_H)\}_{l=0}^{L}\right),$$

and its power spectrum $P_l = \sum_m |Q_{lm}|^2$ gives rotation-invariant fingerprints of the *relative placement and anisotropy* of aspirin's aromatic ring, carboxyl, ester, and methyl groups, capturing chemically meaningful orientation without dependence on global pose.

Example 17 (Hydrogen-bonded acetic acid–water 1:1 complex)**.** *Atom-centric hypergraph.* Let $H = (V_H, E_H, \iota)$ model $CH_3COOH \cdot H_2O$ with a hydrogen bond $\mathrm{H}_{\mathrm{w1}} \cdots \mathrm{O}_{\mathrm{d}}$ (water H to carbonyl O). Atoms:

$$V_H = \{\mathrm{C}_1, \mathrm{C}_2, \mathrm{O}_{\mathrm{d}}, \mathrm{O}_{\mathrm{s}}, \mathrm{H}_{1\mathrm{a}}, \mathrm{H}_{1\mathrm{b}}, \mathrm{H}_{1\mathrm{c}}, \mathrm{H}_{\mathrm{a}}, \mathrm{O}_{\mathrm{w}}, \mathrm{H}_{\mathrm{w1}}, \mathrm{H}_{\mathrm{w2}}\},$$

with coordinates $\mathbf{r}_V$ from, e.g., a cluster optimization. Define chemically meaningful hyperedges:

$$\begin{aligned} &e_{\mathrm{CH}_3} = \{\mathrm{C}_1, \mathrm{H}_{1\mathrm{a}}, \mathrm{H}_{1\mathrm{b}}, \mathrm{H}_{1\mathrm{c}}\}, \qquad e_{\mathrm{C=O}} = \{\mathrm{C}_2, \mathrm{O}_{\mathrm{d}}\}, \\ &e_{\mathrm{OH}} = \{\mathrm{O}_{\mathrm{s}}, \mathrm{H}_{\mathrm{a}}\}, \qquad e_{\mathrm{H_2O}} = \{\mathrm{O}_{\mathrm{w}}, \mathrm{H}_{\mathrm{w1}}, \mathrm{H}_{\mathrm{w2}}\}, \qquad e_{\mathrm{HB}} = \{\mathrm{H}_{\mathrm{w1}}, \mathrm{O}_{\mathrm{d}}\}, \end{aligned}$$

and $E_H = \{e_{\mathrm{CH}_3}, e_{\mathrm{C=O}}, e_{\mathrm{OH}}, e_{\mathrm{H_2O}}, e_{\mathrm{HB}}\}$, with $\iota(e) = e$.

Geometry and weights. Set $w_V(v) := |q(v)|$ from any consistent charge scheme (e.g. AM1-BCC). For each hyperedge e, take

$$\mathbf{r}_E(e) = \frac{1}{|e|} \sum_{v \in e} \mathbf{r}_V(v), \qquad w_E(e) = \sum_{v \in e} w_V(v).$$

Multipoles and carrier. Choose the fused carrier $\mathsf{Car} = V_H \cup E_H$ to encode both atom-level and group-level geometry. For order $L \in \mathbb{N}$ define

$$Q_{lm}\,(H, \mathsf{Car} = V_H \cup E_H) = \sum_{a \in V_H \cup E_H} w(a)\, \|\widehat{\mathbf{r}}(a)\|^l\, Y_{lm}(\theta_a, \phi_a)^*.$$

Thus

$$\mathsf{MpMHG}_L\,(H, V_H \cup E_H) = \left(H, \{Q_l(H, V_H \cup E_H)\}_{l=0}^{L}\right).$$

Interpretation. The invariants $\{P_l\}$ quantify the *directionality and strength* of the hydrogen-bonded dimer: they differentiate contact versus solvent-separated motifs via e_{HB}, capture the orientation between C=O and water dipoles, and summarize the cluster's anisotropy while being invariant to overall rotation/translation.

Theorem 10 (Rigid-motion invariances)**.** *Let $R \in SO(3)$ and $\mathbf{t} \in \mathbb{R}^3$, and let $H^{(R,\mathbf{t})}$ be obtained from H by replacing each position with $R\mathbf{r}(\cdot) + \mathbf{t}$ while keeping $V_H, E_H, \iota, \lambda_\bullet, w_\bullet$ fixed. Then, for any carrier* Car*:*

(i) *(Translation invariance after centering)* $Q_{lm}\left(H^{(I,\mathbf{t})}, \mathsf{Car}\right) = Q_{lm}\,(H, \mathsf{Car})$ *for all* l, m.

(ii) *(Rotation covariance)* $Q_l\left(H^{(R,\mathbf{0})}, \mathsf{Car}\right) = D^{(l)}(R)\, Q_l\,(H, \mathsf{Car})$, *with* $D^{(l)}(R)$ *the unitary Wigner D-matrix. Hence* P_l *is rotation invariant:* $P_l\left(H^{(R,\mathbf{0})}, \mathsf{Car}\right) = P_l\,(H, \mathsf{Car})$.

Proof. Centering removes the translation so (i) follows immediately. The spherical harmonics satisfy $Y_{lm}(R \cdot \Omega) = \sum_{m'} D^{(l)}_{mm'}(R) Y_{lm'}(\Omega)$; substitution yields (ii) and the invariance of $P_l = \|Q_l\|_2^2$ by unitarity. □

Theorem 11 (MpMHG generalizes both MMG and Molecular HyperGraphs)**.** *Let $\mathcal{C}_{\mathrm{MMG}}$ be the class of Multipolar Molecular Graphs (as defined for graphs with geometric atom positions and weights), and let $\mathcal{C}_{\mathrm{MHG}}$ be the class of molecular hypergraphs (with no geometry). Then:*

(a) *(*Embedding of MMG into MpMHG*) Given any MMG of order L built from a molecular* graph $G = (V, E, \lambda_V, \lambda_E, \mathbf{r}, w)$, *construct an* atom-centric *hypergraph H with* $V_H := V$, $E_H := \{\{u, v\} : \{u, v\} \in E\}$ *and* $\iota(\{u, v\}) = \{u, v\}$. *Set* $\mathbf{r}_V = \mathbf{r}$, *choose any* $\mathbf{r}_E$ *(unused),* $w_V = w$, *and arbitrary* w_E. *With carrier* $\mathsf{Car} = V_H$, *the multipoles satisfy*

$$Q_{lm}\,(H, \mathsf{Car} = V_H) = \sum_{v \in V} w(v)\, \|\mathbf{r}(v) - \mathbf{r}_c\|^l\, Y_{lm}(\theta_v, \phi_v)^* = Q_{lm}(G),$$

hence $\mathsf{MpMHG}_L(H, V_H)$ *reproduces the MMG of order L.*

(b) *(*Forgetting map to MHG*) For any MpMHG* $\mathsf{MpMHG}_L(H, \mathsf{Car})$, *the projection*

$$\mathsf{U} : \; (H, \{Q_l\}_{l \le L}) \longmapsto H = (V_H, E_H, \iota, \lambda_V, \lambda_E),$$

forgets coordinates, weights, and multipoles, yielding an object in $\mathcal{C}_{\mathrm{MHG}}$. Thus the underlying hypergraph structure is preserved, i.e. U *is a surjective homomorphism from MpMHG to MHG.*

Consequently, MpMHG strictly *contains both $\mathcal{C}_{\mathrm{MMG}}$ (via (a)) and $\mathcal{C}_{\mathrm{MHG}}$ (via the existence of preimages under* U*), and unifies geometric multipole descriptors with hyperedge incidence.*

Proof. (a) By construction the atom-centric hyperedges have size 2, and $\mathsf{Car} = V_H$ restricts the multipole sum to vertices only, exactly matching the MMG definition (same centroid, positions, and weights), hence Q_{lm} agree term-by-term.

(b) The map U discards $\mathbf{r}_\bullet, w_\bullet$ and $\{Q_l\}$, leaving $(V_H, E_H, \iota, \lambda_\bullet)$ unchanged, which is a valid molecular hypergraph. Surjectivity is immediate: given any molecular hypergraph H_0, choose arbitrary positions and positive weights to obtain some MpMHG whose U-image is H_0. Strict inclusion follows since non-isomorphic MpMHGs can map to the same H_0 (different geometries/weights produce different multipoles while sharing incidence), and not every MHG carries multipole data. □

3.4.3. Multipolar molecular SuperHyperGraph

Multipolar Molecular SuperHyperGraphs further generalize these constructions by embedding multipolar descriptors into hierarchical superhypergraph structures, unifying graphs, hypergraphs, and multi-level superhypergraph frameworks, and enabling a deeper representation of complex molecular systems across multiple structural scales.

Definition 25 (Geometry and weights on $H^{(n)}$)**.** Let $\mathbf{r}_V : V \to \mathbb{R}^3$ and $\mathbf{r}_E : E \to \mathbb{R}^3$ assign positions to vertices and superedges (e.g. atomic coordinates and geometric representatives/barycenters). Let $w_V : V \to \mathbb{R}_{>0}$ and $w_E : E \to \mathbb{R}_{>0}$ be positive weights (e.g. masses/charges/importance). For $a \in V \cup E$ set

$$\mathbf{r}(a) := \begin{cases} \mathbf{r}_V(a), & a \in V, \\ \mathbf{r}_E(a), & a \in E, \end{cases} \qquad w(a) := \begin{cases} w_V(a), & a \in V, \\ w_E(a), & a \in E. \end{cases}$$

Definition 26 (Carrier, centering, and multipoles on a superhypergraph)**.** Let the *carrier* be any of

$$\mathsf{Car} \in \{V,\ E,\ V \cup E\}.$$

Define the weighted centroid and centered positions by

$$W := \sum_{a \in \mathsf{Car}} w(a), \qquad \mathbf{r}_c := W^{-1} \sum_{a \in \mathsf{Car}} w(a)\, \mathbf{r}(a), \qquad \widehat{\mathbf{r}}(a) := \mathbf{r}(a) - \mathbf{r}_c.$$

Let $r_a := \|\widehat{\mathbf{r}}(a)\|$ and (θ_a, ϕ_a) be the spherical angles of $\widehat{\mathbf{r}}(a)$. For each rank $l \geq 0$ and $m \in \{-l, \ldots, l\}$ define the (Condon–Shortley) spherical-harmonic multipoles

$$Q_{lm}\left(H^{(n)}, \mathsf{Car}\right) := \sum_{a \in \mathsf{Car}} w(a)\, r_a^l\, Y_{lm}(\theta_a, \phi_a)^*,$$

and set $Q_l := (Q_{l,-l}, \ldots, Q_{l,l}) \in \mathbb{C}^{2l+1}$ and the *power spectrum* $P_l := \sum_{m=-l}^{l} |Q_{lm}|^2 = \|Q_l\|_2^2$.

Definition 27 (Multipolar Molecular SuperHyperGraph (MpMSHG))**.** Fix an order $L \in \mathbb{N}$ and a carrier $\mathsf{Car} \in \{V, E, V \cup E\}$. A *Multipolar Molecular SuperHyperGraph of order L* is the tuple

$$\mathsf{MpMSHG}_L\left(H^{(n)}, \mathsf{Car}\right) := \left(H^{(n)},\ \mathbf{r}_V, \mathbf{r}_E,\ w_V, w_E,\ \{Q_l(H^{(n)}, \mathsf{Car})\}_{l=0}^{L}\right).$$

When rotation invariance is required, the representation $\{P_l(H^{(n)}, \mathsf{Car})\}_{l=0}^{L}$ is used instead of $\{Q_l\}$.

Example 18 (Protein–ligand binding pocket as an MpMSHG)**.** *System.* Consider a protein pocket formed by residues $\mathcal{R} = \{R_1, \ldots, R_k\}$ and a bound ligand L. Let $A(X)$ denote the set of atoms belonging to entity X (a residue or the ligand). Let $B := \bigcup_{i=1}^{k} A(R_i) \cup A(L)$ be the base set of atoms with 3D coordinates $\{\mathbf{r}(a) \in \mathbb{R}^3 : a \in B\}$ obtained from a PDB structure, and let $\{q(a)\}$ be atomic partial charges.

n–SuperHyperGraph (n = 2). Define rank–1 vertices (residue and ligand "superatoms")

$$v_{R_i} := A(R_i) \in \mathcal{P}^1(B) \quad (i = 1, \ldots, k), \qquad v_L := A(L) \in \mathcal{P}^1(B).$$

Set $V := \{v_{R_1}, \ldots, v_{R_k}, v_L\} \subseteq \mathcal{P}^{\leq 2}(B)$. Define the rank–2 "binding–site" superedge

$$e_{\text{site}} := \{\, v_{R_1}, \ldots, v_{R_k}, v_L \,\} \in \mathcal{P}(V)\ (= \mathcal{P}^2(B)).$$

Hence $E := \{e_{\text{site}}\}$ and the hierarchy constraint $\mathrm{rk}(u) < \mathrm{rk}(e)$ holds for all $u \in e \in E$.

Geometry and weights. For a rank–1 vertex v_R (a residue),

$$\mathbf{r}_V(v_R) := \frac{1}{|A(R)|} \sum_{a \in A(R)} \mathbf{r}(a), \qquad w_V(v_R) := \sum_{a \in A(R)} |q(a)|.$$

For the ligand v_L use the same formulas with $R \leftarrow L$. For the rank–2 superedge

$$\mathbf{r}_E(e_{\text{site}}) := \frac{1}{k+1} \sum_{u \in e_{\text{site}}} \mathbf{r}_V(u), \qquad w_E(e_{\text{site}}) := \sum_{u \in e_{\text{site}}} w_V(u).$$

Multipoles and carrier. Choose carrier $\mathsf{Car} = E$ (superedge level) and order $L \geq 0$. Compute the weighted centroid $\mathbf{r}_c$ over Car, the centered positions $\widehat{\mathbf{r}}(e)$, and for $l \geq 0,\ m = -l, \ldots, l$ set

$$Q_{lm}\left(H^{(2)}, \mathsf{Car} = E\right) := \sum_{e \in E} w_E(e)\, \|\widehat{\mathbf{r}}(e)\|^l\, Y_{lm}(\theta_e, \phi_e)^*.$$

The MpMSHG is

$$\mathsf{MpMSHG}_L\left(H^{(2)}, E\right) = \left(H^{(2)}, \mathbf{r}_V, \mathbf{r}_E, w_V, w_E, \{Q_l\}_{l=0}^{L}\right).$$

The spectrum $\{P_l = \sum_m |Q_{lm}|^2\}$ summarizes the pocket's anisotropy and the ligand's placement *independently of overall rotation/translation*, reflecting, e.g., how "directional" the binding environment is around the ligand.

Example 19 (Lithium–ion first–shell solvation complex in a battery electrolyte)**.** *System.* Consider a Li^+ cation solvated by carbonate solvents $\mathcal{S} = \{S_1, \ldots, S_r\}$ (e.g. EC/EMC) and the anion A^- (e.g. PF_6^-). Let $A(X)$ be the atom set of species $X \in \{\mathrm{Li}^+\} \cup \mathcal{S} \cup \{A^-\}$, with Cartesian coordinates $\mathbf{r}(a)$ from an MD snapshot, and partial charges $q(a)$.

n–SuperHyperGraph (n = 2). Define rank–1 vertices for each chemical species:

$$v_{\mathrm{Li}} := A(\mathrm{Li}^+), \quad v_{S_j} := A(S_j)\ (j = 1, \ldots, r), \quad v_A := A(A^-).$$

Let $V := \{v_{\mathrm{Li}}, v_{S_1}, \ldots, v_{S_r}, v_A\}$. Let the *first–solvation–shell* around Li^+ be the set

$$\mathsf{Shell}_1 := \left\{ v_{S_j} : \min_{a \in A(S_j)} \|\mathbf{r}(a) - \mathbf{r}(\mathrm{Li})\| \leq \rho_c \right\},$$

for a coordination cutoff $\rho_c > 0$. Define rank–2 superedges

$$e_{\mathrm{CIP}} := \{v_{\mathrm{Li}}, v_A\}, \qquad e_{\mathrm{Solv}} := \{v_{\mathrm{Li}}\} \cup \mathsf{Shell}_1, \qquad E := \{e_{\mathrm{CIP}}, e_{\mathrm{Solv}}\}.$$

The hierarchy constraint holds since v's have rank 1 and e's rank 2.

Geometry and weights. For a vertex v_X (species X),

$$\mathbf{r}_V(v_X) := \frac{1}{|A(X)|} \sum_{a \in A(X)} \mathbf{r}(a), \qquad w_V(v_X) := \sum_{a \in A(X)} |q(a)|.$$

For a superedge e,

$$\mathbf{r}_E(e) := \frac{1}{|e|} \sum_{u \in e} \mathbf{r}_V(u), \qquad w_E(e) := \sum_{u \in e} w_V(u).$$

Multipoles and carrier. Choose carrier $\mathsf{Car} = V \cup E$ to fuse molecular (vertex) and aggregate (superedge) geometry. For order $L \geq 0$,

$$Q_{lm}\left(H^{(2)}, \mathsf{Car} = V \cup E\right) := \sum_{a \in V \cup E} w(a)\, \|\widehat{\mathbf{r}}(a)\|^l\, Y_{lm}(\theta_a, \phi_a)^*.$$

Thus

$$\mathsf{MpMSHG}_L\left(H^{(2)}, V \cup E\right) = \left(H^{(2)}, \mathbf{r}_V, \mathbf{r}_E, w_V, w_E, \{Q_l\}_{l=0}^{L}\right).$$

The rotationally invariant powers $\{P_l\}$ capture the *shape and directionality* of Li^+ coordination (tetra-/penta-/hexa–coordination anisotropy) and the presence of a contact–ion pair (e_{CIP}) versus solvent–separated structures (e_{Solv}), all robust to global motions of the simulation box.

Theorem 12 (Rigid-motion invariances on superhypergraphs)**.** *Let $R \in SO(3)$ and $\mathbf{t} \in \mathbb{R}^3$. Denote by $H^{(n)}_{(R,\mathbf{t})}$ the object obtained by replacing every position $\mathbf{r}(\cdot)$ with $R\mathbf{r}(\cdot) + \mathbf{t}$ while keeping $V, E, \ell_\bullet, w_\bullet$ unchanged. Then for any carrier* Car *and every l:*

(1) Translation invariance after centering: $Q_{lm}\left(H^{(n)}_{(I,\mathbf{t})}, \mathsf{Car}\right) = Q_{lm}\left(H^{(n)}, \mathsf{Car}\right)$.

(2) Rotation covariance: $Q_l\left(H^{(n)}_{(R,\mathbf{0})}, \mathsf{Car}\right) = D^{(l)}(R)\, Q_l\left(H^{(n)}, \mathsf{Car}\right)$, *with* $D^{(l)}(R)$ *the unitary Wigner D-matrix. In particular* P_l *is rotation invariant.*

Proof. Centering removes any global translation, proving (1). Spherical harmonics obey $Y_{lm}(R \cdot \Omega) = \sum_{m'=-l}^{l} D^{(l)}_{mm'}(R)\, Y_{lm'}(\Omega)$; substituting into the definition of Q_{lm} yields (2). Unitarity of $D^{(l)}$ implies $\|Q_l\|_2$ and hence P_l is preserved. □

Theorem 13 (MpMSHG generalizes MMG, MpMHG, and MSHG)**.** *Let* $\mathcal{C}_{\mathrm{MMG}}$ *be the class of* Multipolar Molecular Graphs *(built on graphs with atomic positions/weights),* $\mathcal{C}_{\mathrm{MpMHG}}$ *the class of* Multipolar Molecular HyperGraphs*, and* $\mathcal{C}_{\mathrm{MSHG}}$ *the class of (unweighted, geometry-free)* Molecular SuperHyperGraphs*. Then:*

(a) (Embedding of MMG) *For any MMG of order* L *on a molecular graph* $G = (V_G, E_G, \mathbf{r}, w)$*, define an* $n = 1$ *superhypergraph* $H^{(1)}$ *by* $V := V_G$ *(rank 0),* $E := \{\{u, v\} : \{u, v\} \in E_G\}$ *(rank 1), and incidence* $u, v \in \{u, v\}$*. Set* $\mathbf{r}_V = \mathbf{r}$*, choose arbitrary* $\mathbf{r}_E$*,* $w_V = w$*, arbitrary* w_E*, and take* $\mathsf{Car} = V$*. Then for all* l, m

$$Q_{lm}\left(H^{(1)}, \mathsf{Car} = V\right) = \sum_{v \in V_G} w(v)\, \|\mathbf{r}(v) - \mathbf{r}_c\|^l\, Y_{lm}(\theta_v, \phi_v)^* = Q_{lm}(G),$$

so $\mathsf{MpMSHG}_L(H^{(1)}, V)$ *reproduces the MMG of order* L*.*

(b) (Embedding of MpMHG) *Given any MpMHG of order* L *on a molecular hypergraph*

$$H = (V_H, E_H, \iota, \mathbf{r}_V, \mathbf{r}_E, w_V, w_E),$$

form $H^{(1)}$ *by* $V := V_H$ *(rank 0) and* $E := \{\iota(e) : e \in E_H\}$ *(rank 1). Keeping the same* $\mathbf{r}_V, \mathbf{r}_E, w_V, w_E$ *and the same choice of* $\mathsf{Car} \in \{V, E, V \cup E\}$*, the multipole sums are identical term-by-term; hence* $\mathsf{MpMSHG}_L(H^{(1)}, \mathsf{Car})$ *reproduces the given MpMHG.*

(c) (Forgetting map to MSHG) *For any MpMSHG, the projection*

$$\mathsf{U}: \left(H^{(n)}, \mathbf{r}_\bullet, w_\bullet, \{Q_l\}_{l \le L}\right) \longmapsto H^{(n)},$$

forgets geometry, weights, and multipoles, yielding an element of $\mathcal{C}_{\mathrm{MSHG}}$*. Conversely, every* $H^{(n)}_0 \in \mathcal{C}_{\mathrm{MSHG}}$ *is the image of some MpMSHG obtained by assigning any positive weights and any positions in* $\mathbb{R}^3$ *(e.g. distinct barycentric placements) and then computing* $\{Q_l\}$*.*

Therefore $\mathcal{C}_{\mathrm{MMG}}$ *and* $\mathcal{C}_{\mathrm{MpMHG}}$ *embed into MpMSHGs, and* $\mathcal{C}_{\mathrm{MSHG}}$ *is a quotient under* U*; the containments are strict.*

Proof. (a) The construction makes every graph edge a size–2 superedge (rank 1), so the hierarchy holds. Choosing $\mathsf{Car} = V$ restricts the multipole sum to vertices; centroids, radii, and weights coincide with those of G, hence the displayed equality holds for each l, m.

(b) By definition $E = \{\iota(e)\}$ and $V = V_H$, so the incidence structure is preserved and every summand $w(\cdot)\, r^l Y_{lm}(\cdot)$ in the MpMHG sum appears identically in the MpMSHG sum for the same carrier; thus the multipoles agree componentwise.

(c) The map U leaves $(V, E, \ell_\bullet)$ unchanged, producing a valid MSHG. Surjectivity: given $H^{(n)}_0$, pick any injective placement of $V \cup E$ in $\mathbb{R}^3$ and any positive weights to obtain an MpMSHG that projects to $H^{(n)}_0$. Strictness: two MpMSHGs with the same (V, E) but different geometries/weights yield different multipoles yet have the same image under U; likewise an MSHG generally lacks the additional geometric/multipolar data and cannot reconstruct them uniquely. □

4. Conclusion

In this paper, we investigated Weighted, Rough, Neural, and Multipolar frameworks for Molecular Graphs, Molecular HyperGraphs, and Molecular SuperHyperGraphs. More specifically, we described how weighted structures can encode chemically meaningful quantities such as bond strengths, traversal costs, and atomic importance; how rough structures can model indiscernibility and coarse observational uncertainty; how neural extensions can support message passing and molecular property prediction on graph, hypergraph, and superhypergraph domains; and how multipolar frameworks can incorporate geometric and directional information in a mathematically structured manner. Through these viewpoints, the paper provided a unified perspective for representing molecular systems at ordinary, higher-order, and hierarchical levels.

In future work, it would be natural to extend the graph concepts studied here by incorporating additional uncertainty-aware and direction-sensitive frameworks, including Fuzzy Graphs [62], Neutrosophic Graphs [63,64], Bidirected Graphs [65], Quadripartitioned Neutrosophic Graphs [66], Pentapartitioned Neutrosophic Graphs [67,68], and Plithogenic Graphs [69]. Such extensions may lead to richer molecular models capable of simultaneously handling hierarchy, ambiguity, polarity, multi-source uncertainty, and directional interactions. They may also provide useful mathematical foundations for future applications in molecular representation learning, chemical biology, reaction-network analysis, and decision-support problems involving complex chemical systems.

Use of Generative AI and AI-Assisted Tools

We used generative AI and AI-assisted tools only for limited editorial support, such as English grammar checking, and did not use them in any manner that violates ethical standards

Disclaimer

This work presents theoretical concepts that have not yet undergone practical testing or validation. Future researchers are encouraged to apply and assess these ideas in empirical contexts. While every effort has been made to ensure accuracy and appropriate referencing, unintentional errors or omissions may still exist. Readers are advised to verify referenced materials on their own. The views and conclusions expressed here are the authors' own and do not necessarily reflect those of their affiliated organizations.

Funding Information: This study did not receive any financial or external support from organizations or individuals.

Data Availability: This research is purely theoretical, involving no data collection or analysis. We encourage future researchers to pursue empirical investigations to further develop and validate the concepts introduced here.

Acknowledgments: We extend our sincere gratitude to everyone who provided insights, inspiration, and assistance throughout this research. We particularly thank our readers for their interest and acknowledge the authors of the cited works for laying the foundation that made our study possible. We also appreciate the support from individuals and institutions that provided the resources and infrastructure needed to produce and share this paper. Finally, we are grateful to all those who supported us in various ways during this project.

Conflicts of Interest: The author confirms that there are no conflicts of interest related to the research or its publication.

References

[1] Gross, J. L., Yellen, J., & Anderson, M. (2018). *Graph Theory and Its Applications*. Chapman and Hall/CRC.

[2] Bretto, A. (2013). *Hypergraph Theory: An Introduction*. Springer.

[3] Cai, D., Song, M., Sun, C., Zhang, B., Hong, S., & Li, H. (2022). Hypergraph structure learning for hypergraph neural networks. In *Proceedings of IJCAI* (pp. 1923–1929).

[4] Gao, Y., Zhang, Z., Lin, H., Zhao, X., Du, S., & Zou, C. (2020). Hypergraph learning: Methods and practices. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(5), 2548–2566.

[5] Hamidi, M., Smarandache, F., & Davneshvar, E. (2022). Spectrum of superhypergraphs via flows. *Journal of Mathematics, 2022*(1), 195812.

[6] Hamidi, M., & Taghinezhad, M. (2023). *Application of Superhypergraphs-Based Domination Number in Real World*. Infinite Study.

[7] Fujita, T., & Smarandache, F. (2026). *Hypergraph and Superhypergraph Theory With Applications*. Neutrosophic Science International Association (NSIA) Publishing House.

[8] Fujita, T., & Smarandache, F. (2026). *Representing Higher-Order Networks: A Survey of Graph-Based Frameworks*. Neutrosophic Science International Association (NSIA) Publishing House.

[9] Roshdy, E., Khashaba, M., & Ahmed Ali, M. E. (2025). Neutrosophic super-hypergraph fusion for proactive cyberattack countermeasures: A soft computing framework. *Neutrosophic Sets and Systems, 94*, 232–252.

[10] Diestel, R. (2024). *Graph theory*. Springer.

[11] Smarandache, F. (2020). *Extension of Hypergraph to N-Superhypergraph, and to Plithogenic N-Superhypergraph, and Extension of Hyperalgebra to N-Ary Classical-/Neutro-/Anti-Hyperalgebra*. Infinite Study.

[12] Fujita, T. (2025). Molecular fuzzy graphs, hypergraphs, and superhypergraphs. *Journal of Intelligent Decision and Computational Modelling*, *1*(3), 158–171.

[13] Hamidi, M., Smarandache, F., & Taghinezhad, M. (2023). *Decision Making Based on Valued Fuzzy Superhypergraphs*. Infinite Study.

[14] Alqahtani, M. (2025). Intuitionistic fuzzy quasi-supergraph integration for social network decision making. *International Journal of Analysis and Applications*, *23*, 137–137.

[15] Fujita, T., Saqlain, M., & Gulistan, M. (2026). Hierarchical network modeling with intuitionistic fuzzy superhypergraphs. *Multicriteria Algorithms with Applications*, *10*(1), 1–22.

[16] Fujita, T. (2024). Vague superhypergraph. *Uncertainty Discourse and Applications*.

[17] Berrocal Villegas, S. M., Fritas, W. M., Berrocal Villegas, C. R., Flores Fuentes Rivera, M. Y., Espejo Rivera, R., Bautista Puma, L. D., & Macazana Fernández, D. M. (2025). Using plithogenic n-superhypergraphs to assess the degree of relationship between formation skills and digital competencies. *Neutrosophic Sets and Systems*, *84*(1), 41.

[18] Fujita, T., & Smarandache, F. (2025). Soft directed n-superhypergraphs with some real-world applications. *European Journal of Pure and Applied Mathematics*, *18*(4), 6643–6643.

[19] Fujita, T. (2024). Review of some superhypergraph classes: Directed, bidirected, soft, and rough. In *Advancing Uncertain Combinatorics Through Graphization, Hyperization, and Uncertainization: Fuzzy, Neutrosophic, Soft, Rough, and Beyond* (Vol. 2).

[20] Fujita, T. (2025). Modeling complex hierarchical systems with weighted and signed superhypergraphs: Foundations and applications. *Open Journal of Discrete Applied Mathematics (ODAM)*, *8*(3), 20–39.

[21] Du, W., Zhang, S., Cai, Z., Li, X., Liu, Z., Fang, J., Wang, J., Wang, X., & Wang, Y. (2025). Molecular merged hypergraph neural network for explainable solvation Gibbs free energy prediction. *Research*, *8*, 0740.

[22] Alawad, M. A., Al Araji, H. N., Singh, N. S. S., & Hussein, A. A. (2025). Hamnet: Hierarchical attention-enhanced molecular hypergraph network for medication recommendation with data imbalance mitigation. *Egyptian Informatics Journal*, *32*, 100849.

[23] Tkachenko, A. V. (2026). Structural and compositional complexities of hierarchical self-assembly: A hypergraph approach. *The Journal of Chemical Physics*, *164*(4).

[24] Tu, Z., Fan, W., & Zhang, Y. (2025). Hyperloc: A hypergraph-enhanced multimodal framework for multi-label RNA subcellular localization prediction. In *2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 190–195). IEEE.

[25] Fujita, T. (2025). Multi-superhypergraph neural networks: A generalization of multi-hypergraph neural networks. *Neutrosophic Computing and Machine Learning*, *39*, 328–347.

[26] Zhu, J., Zhao, X., Hu, H., & Gao, Y. (2019). Emotion recognition from physiological signals using multi-hypergraph neural networks. In *2019 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 610–615). IEEE.

[27] Feng, Y., You, H., Zhang, Z., Ji, R., & Gao, Y. (2019). Hypergraph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. *33*, pp. 3558–3565).

[28] Fujita, T., & Smarandache, F. (2026). *Theoretical Foundations of Superhypergraph and Plithogenic Graph Neural Networks*. Neutrosophic Science International Association (NSIA) Publishing House.

[29] Liu, S., Chen, M., Yao, X., & Liu, H. (2025). Fingerprint-enhanced hierarchical molecular graph neural networks for property prediction. *Journal of Pharmaceutical Analysis*, 101242.

[30] Ekström Kelvinius, F., Georgiev, D., Toshev, A., & Gasteiger, J. (2023). Accelerating molecular graph neural networks via knowledge distillation. *Advances in Neural Information Processing Systems*, *36*, 25761–25792.

[31] Vik, D., Pii, D., Mudaliar, C., Nørregaard-Madsen, M., & Kontijevskis, A. (2024). Performance and robustness of small molecule retention time prediction with molecular graph neural networks in industrial drug discovery campaigns. *Scientific Reports*, *14*(1), 8733.

[32] Chen, J., & Schwaller, P. (2024). Molecular hypergraph neural networks. *The Journal of Chemical Physics*, *160*(14), 144307.

[33] Casetti, N., Nevatia, P., Chen, J., Schwaller, P., & Coley, C. W. (2024). Comment on "Molecular hypergraph neural networks" [J. Chem. Phys. *160*, 144307 (2024)]. *The Journal of Chemical Physics*, *161*(20).

[34] Das, A. K., Das, R., Das, S., Debnath, B. K., Granados, C., Shil, B., & Das, R. (2025). A comprehensive study of neutrosophic superhyper BCI-semigroups and their algebraic significance. *Transactions on Fuzzy Sets and Systems*, *8*(2), 80.

[35] Smarandache, F. (2023). Foundation of the superhypersoft set and the fuzzy extension superhypersoft set: A new vision. *Neutrosophic Systems with Applications*, *11*, 48–51.

[36] Jech, T. (2003). *Set Theory: The Third Millennium Edition, Revised and Expanded*. Springer.
[37] Smarandache, F. (2024). Foundation of superhyperstructure & neutrosophic superhyperstructure. *Neutrosophic Sets and Systems, 63*(1), 21.
[38] Berge, C. (1984). *Hypergraphs: Combinatorics of Finite Sets* (Vol. 45). Elsevier.
[39] Vizuete, G. X., Gallardo, C. F., & Vizuete, G. (2025). Structured analysis of a generator set lubrication system using vertex articulation in plithogenic n-superhypergraphs. *Neutrosophic Sets and Systems*, *89*, 3–103.
[40] Fujita, T., & Smarandache, F. (2025). Competition super-hypergraphs: Revealing hierarchical competition in real-world networks. *Journal of Algebra and Applied Mathematics, 23*(2), 97–116.
[41] Akram, M., & Shahzadi, G. (2018). Hypergraphs in m-polar fuzzy environment. *Mathematics, 6*(2), 28.
[42] Fujita, T., & Smarandache, F. (2024). A concise study of some superhypergraph classes. *Neutrosophic Sets and Systems, 77*, 548–593.
[43] Sebastian, L. (2021). Topological indices of molecular graphs of some anti-cancer drugs. *SGS-Engineering & Sciences, 1*(01).
[44] Goupil, B., Joly, A., Pancino, N., Bongini, P., Scarselli, F., & Bianchini, M. (2025). MolGMP: A Markov approach for molecular graph generation with GNNs. *Neurocomputing*, 131066.
[45] Li, C., Xiang, H., Du, W., Ma, T., Chen, H., Zeng, X., & Xu, L. (2025). GraphGIM: Rethinking molecular graph contrastive learning via geometry image modeling. *BMC Biology, 23*(1), 189.
[46] Rahman, A., Poirel, C. L., Badger, D. J., & Murali, T. M. (2012). Reverse engineering molecular hypergraphs. In *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine* (pp. 68–75).
[47] Kajino, H. (2018). Molecular hypergraph grammar with its application to molecular optimization. *arXiv*, abs/1809.02745.
[48] Fujita, T. (2025). Exploration of graph classes and concepts for superhypergraphs and n-th power mathematical structures. *Advancing Uncertain Combinatorics through Graphization, Hyperization, and Uncertainization: Fuzzy, Neutrosophic, Soft, Rough, and Beyond, 3*(4), 512.
[49] Ivanciuc, O. (2000). QSAR comparative study of Wiener descriptors for weighted molecular graphs. *Journal of Chemical Information and Computer Sciences, 40*(6), 1412–1422.
[50] Sorgun, S., & Birgin, K. (2025). Vertex-edge-weighted molecular graphs: A study on topological indices and their relevance to physicochemical properties of drugs used in cancer treatment. *Journal of Chemical Information and Modeling, 65*(4), 2093–2106.
[51] Berinde, Z. (2006). Vertex- and edge-weighted molecular graphs for amines. *Revue Roumaine de Chimie, 51*(11), 1131.
[52] Pawlak, Z. (1982). Rough sets. *International Journal of Computer & Information Sciences, 11*, 341–356.
[53] Broumi, S., Smarandache, F., & Dhar, M. (2014). Rough neutrosophic sets. *Infinite Study, 32*, 493–502.
[54] Akram, M., Arshad, M., & Shumaiza. (2018). Fuzzy rough graph theory with applications. *International Journal of Computational Intelligence Systems, 12*, 90–107.
[55] Patrascu, V. (2007). Rough sets on four-valued fuzzy approximation space. In *2007 IEEE International Fuzzy Systems Conference* (pp. 1–5). IEEE.
[56] Pawlak, Z., Polkowski, L., & Skowron, A. (2001). Rough set theory. *KI, 15*(3), 38–39.
[57] Pawlak, Z., & Skowron, A. (2007). Rudiments of rough sets. *Information Sciences, 177*(1), 3–27.
[58] He, T., Chen, Y., & Shi K. (2006). Weighted rough graph and its application. In *Sixth International Conference on Intelligent Systems Design and Applications, 1*, 486–491.
[59] Tong, H., Peijun, X., & Kaiquan, S. (2008). Application of rough graph in relationship mining. *Journal of Systems Engineering and Electronics, 19*(4), 742–747.
[60] Noor, R., Irshad, L., & Javaid, I. (2017). Soft rough graphs. *arXiv preprint arXiv:1707.05837*.
[61] Colby, S. M., Shapiro, M. R., Lin, A., Bilbao, A., Broeckling, C. D., Purvine, E., & Joslyn, C. A. (2024). Introducing molecular hypernetworks for discovery in multidimensional metabolomics data. *Journal of Proteome Research*, *23*(11), 4789–4801.
[62] Al-Hawary, T. (2011). Complete fuzzy graphs. *International Journal of Mathematical Combinatorics, 4*, 26.
[63] Broumi, S., Talea, M., Bakali, A., & Smarandache, F. (2016). Interval valued neutrosophic graphs. *Critical Review, XII*, 5–33.
[64] Broumi, S., Talea, M., Bakali, A., Smarandache, F., & Kumar, P. K. (2017). Shortest path problem on single valued neutrosophic graphs. In *2017 International Symposium on Networks, Computers and Communications (ISNCC)* (pp. 1–6). IEEE.
[65] Xu, R., & Zhang, C.-Q. (2005). On flows in bidirected graphs. *Discrete Mathematics, 299*(1–3), 335–343.
[66] Shi, X., Kosari, S., Rashmanlou, H., Broumi, S., & Hussain, S. S. (2023). Properties of interval-valued quadripartitioned neutrosophic graphs with real-life application. *Journal of Intelligent & Fuzzy Systems, 44*(5), 7683–7697.

[67] Das, S., Das, R., & Pramanik, S. (2022). Single valued pentapartitioned neutrosophic graphs. *Neutrosophic Sets and Systems*, *50*(1), 225–238.

[68] Broumi, S., Ajay, D., Chellamani, P., Malayalan, L., Talea, M., Bakali, A., Schweizer, P., & Jafari, S. (2022). Interval valued pentapartitioned neutrosophic graphs with an application to MCDM. *Operational Research in Engineering Sciences: Theory and Applications*, *5*(3), 68–91.

[69] Sultana, F., Gulistan, M., Ali, M., Yaqoob, N., Khan, M., Rashid, T., & Ahmed, T. (2023). A study of plithogenic graphs: Applications in spreading coronavirus disease (COVID-19) globally. *Journal of Ambient Intelligence and Humanized Computing*, *14*(10), 13139–13159.