1.1 Constructing Target Chemical Graphs

This section describes how to construct a target chemical graph in Stages 4 and 5.

1.1.1 Formulating an MILP for a prediction function in Stage 4

In Stage 3, we construct a prediction function $\eta_N : \mathbb{R}^K \to \mathbb{R}$. It is known that the computation process of $\eta_N(x)$ from a vector $x^* \in \mathbb{R}^K$ can be formulated as an MILP with the following property.

**Theorem 1.** ([1, 2]) Let $\mathcal{N}$ be an ANN with a piecewise-linear activation function for an input vector $x \in \mathbb{R}^K$, $n_A$ denote the number of nodes in the architecture and $n_B$ denote the total number of breakpoints over all activation functions. Then there is an MILP $\mathcal{M}(x, y; C_1)$ that consists of variable vectors $x^* \in \mathbb{R}^K$, $y \in \mathbb{R}$, and an auxiliary variable vector $z \in \mathbb{R}^p$ for some integer $p = O(n_A + n_B)$ and a set $C_1$ of $O(n_A + n_B)$ constraints on these variables such that: $\eta_N(x^*) = y^*$ if and only if there is a vector $(x^*, y^*)$ feasible to $\mathcal{M}(x, y; C_1)$.

Solving this MILP delivers a vector $x^* \in \mathbb{R}^K$ such that $\eta_N(x^*) = y^*$ for a target value $y^*$. However, the resulting vector $x^*$ may not admit a chemical graph $G^*$ such that $f(G^*) = x^*$. To ensure that such chemical graph always exists in Stage 4, we further introduce some more constraints for a set of new variables in the next section.

1.1.2 Formulating an MILP for a feature vector and a target specification in Stage 4

In this section, we show an outline of formulation of an MILP that represents the computation process of a feature function $f(G)$ from a chemical graph $G$ and a construction of a target chemical graph $G \in \mathcal{G}(G_C, \sigma_{\text{int}}, \sigma_{\text{ex}})$. Recall that the number of vertices in a target chemical graph is bounded by an upper bound $n^*$ in a specification $(G_C, \sigma_{\text{int}}, \sigma_{\text{ex}})$. However, if we introduce a set of $(n^*)^2$ variables for all pairs of $n^*$ vertices to present all possible graphs for a target chemical graph, then the resulting MILP formulation is hard to solve for $n^* > 20$ due to a larger number of variables and constraints. To overcome this, a sparse representation of chemical graphs has been proposed in the previous applications of the framework for acyclic graphs [3] and $p$-lean graphs [4]. We also define a similar sparse representation to formulate an MILP for our two-layered model.
**Scheme Graphs** We first regard a given seed graph $G_C$ as a digraph and then add some more vertices and edges to construct a digraph, called a scheme graph $SG = (V, E)$ so that any ($\sigma_{int}, \sigma_{ce}$)-extension $H$ of $G_C$ can be chosen as a subgraph of $SG$.

For a given target specification $(G_C, \sigma_{int}, \sigma_{ce})$, define integers that determine the size of a scheme graph $SG$ as follows. $m_C := |E_C|$, $t_C := |V_C|$, $t_T := n_{UB}^{int} - |V_C|$, and $t_F := n^* - n_{LB}^{int}$.

Formally the scheme graph $SG = (V, E)$ is defined with a vertex set $V = V_C \cup V_T \cup V_F$ and an edge set $E = E_C \cup E_T \cup E_F \cup E_{CT} \cup E_{CF} \cup E_{TF}$ that consist of the following sets. See Figure 1 for an illustration of these sets.

**Construction of a $\sigma_{int}$-extension $H^*$ of $G_C$:** Denote the vertex set $V_C$ and the edge set $E_C$ in the seed graph $G_C$ by $V_C = \{v^C_i \mid i \in [1, t_C]\}$ and $E_C = \{e_i \mid i \in [1, m_C]\}$, respectively, where $V_C$ is always included in $H^*$. For including additional interior-vertices in $H^*$, introduce a path $P_T = (V_T = \{v^{T_1}, v^{T_2}, \ldots, v^{T_{t_T}}\}, E_T = \{e^{T_2}, e^{T_3}, \ldots, e^{T_{t_T}}\})$ of length $t_T - 1$ and a set $E_{CT}$ (resp., $E_{TC}$) of directed edges $e^{CT}_{ij} = (v^{C_i}, v^{C_j})$ (resp., $e^{TC}_{ij} = (v^{T_i}, v^{C_j})$) $i \in [1, t_C], j \in [1, t_T]$. In $H^*$, an edge $e_k = (v^{C_i}, v^{C_j}) \in E_{CT} \cup E_{TC}$ is allowed to be replaced with a pure path $P_k$ from vertex $v^{C_i}$ to vertex $v^{C_j}$ that visits a set of consecutive vertices $v^{T_{j+1}}, \ldots, v^{T_{j+p}} \in V_T$ and edge $e^{TC}_{ij} = (v^{C_i}, v^{C_j}) \in E_{CT}$, then edges $e^{T_{j+1}}, e^{T_{j+2}}, \ldots, e^{T_{j+p}} \in E_T$ and finally edge $e^{TC}_{i',j+p} = (v^{T_{j+p}}, v^{C_j}) \in E_{TC}$. The vertices in $V_T$ selected in the path will be vertices in $H^*$.

**Appending leaf paths with additional interior-edges in a ($\sigma_{int}, \sigma_{ce}$)-extension $H$ of $G_C$:** Introduce a path $P_F = (V_F = \{v^{F_1}, v^{F_2}, \ldots, v^{F_{t_F}}\}, E_F = \{e^{F_2}, e^{F_3}, \ldots, e^{F_{t_F}}\})$ of length $t_F - 1$, a set $E_{CF}$ of directed edges $e^{CF}_{i,j} = (v^{C_i}, v^{F_j}), i \in [1, t_C], j \in [1, t_F]$, and a set $E_{TF}$ of directed edges $e^{TF}_{i,j} = (v^{T_i}, v^{F_j}), i \in [1, t_T], j \in [1, t_F]$. In $H$, a leaf path $Q$ with interior-edges that starts from a vertex $v^{C_i} \in V_C$ (resp., $v^{T_i} \in V_T$) visits a set of consecutive vertices $v^{F_j}, v^{F_{j+1}}, \ldots, v^{F_{j+p}} \in V_F$ and edge $e^{CF}_{i,j} = (v^{C_i}, v^{F_j}) \in E_{CF}$ (resp., $e^{TF}_{i,j} = (v^{T_i}, v^{F_j}) \in E_{TF}$) and edges $e^{F_{j+1}}, e^{F_{j+2}}, \ldots, e^{F_{j+p}} \in E_F$. In $H$, the edges and the vertices selected in the path $Q$ are regarded as interior-edges and interior-vertices, respectively.

**Construction of $\rho$-fringe-trees in a ($\sigma_{int}, \sigma_{ce}$)-extension $G$ of $G_C$:** In $H$, the root of a $\rho$-fringe-tree can be any vertex in $V_C \cup V_T \cup V_F$. For each vertex $v = v^{C_i}$ (resp., $v = v^{T_i}$ or $v^{F_i}$), we choose a chemical rooted tree $T$ from the specified set $\mathcal{F}(v)$ (resp., $\mathcal{F}_E$).
Recall that the dimension $K$ of a feature vector $x = f(G)$ used in constructing a prediction function $\eta_C$ over a set of chemical graphs $G$ is $K = 17 + |\Lambda^{\text{int}}(D_\pi)| + |\Lambda^{\text{ex}}(D_\pi)| + |\Gamma^{\text{int}}(D_\pi)| + |\mathcal{F}(D_\pi)|$.

For a target specification $(G_C, \sigma_{\text{int}}, \sigma_{\text{ce}})$, let $\mathcal{F}^*$ denote the set of chemical rooted trees $\psi$ in the sets $\mathcal{F}(v), v \in V_C$ and $\mathcal{F}_E$ and $K^* := 17 + |\Lambda^{\text{int}}(D_\pi)| + |\Lambda^{\text{ex}}(D_\pi)| + |\Gamma^{\text{int}}(D_\pi)| + |\mathcal{F}^*|$. Based on the scheme graph $SG$, we obtain the following MILP formulation $\mathcal{M}(x; g; C_2)$.

**Theorem 2.** Let $(G_C, \sigma_{\text{int}}, \sigma_{\text{ce}})$ be a target specification and $\varphi^* = |\Lambda^{\text{int}}(D_\pi)| + |\Lambda^{\text{ex}}(D_\pi)| + |\Gamma^{\text{int}}(D_\pi)| + |\mathcal{F}^*|$ for sets of chemical elements, edge-configurations and fringe-configurations in $\sigma_{\text{ce}}$. Then there is an MILP $\mathcal{M}(x, g; C_2)$ that consists of variable vectors $x \in \mathbb{R}^{K^*}$ and $g \in \mathbb{R}^q$ for an integer $q = O(n_{\text{UB}}^{\text{int}}(|E_C| + n^*) + (|E_C| + |V|)\varphi^*)$ and a set $C_2$ of $O(|n_{\text{UB}}^{\text{int}}(|E_C| + n^*) + |V|)\varphi^*)$ constraints on $x$ and $g$ such that: $(x^*, g^*)$ is feasible to $\mathcal{M}(x, g; C_2)$ if and only if $g^*$ forms a chemical graph $G \in \mathcal{G}(G_C, \sigma_{\text{int}}, \sigma_{\text{ce}})$ such that $f(G) = x^*$.

Note that our MILP requires only $O(n^*)$ variables and constraints when the branch-parameter $\rho$, integers $|E_C|$, $n_{\text{UB}}^{\text{int}}$ and $\varphi^*$ are constant. We explain the basic idea of our MILP that satisfies Theorem 2. The MILP mainly consists of the following three types of constraints.

- **C1.** Constraints for selecting an underlying graph $H$ of a chemical graph $G \in \mathcal{G}(G_C, \sigma_{\text{int}}, \sigma_{\text{ce}})$ as a subgraph of the scheme graph $SG$;

- **C2.** Constraints for assigning chemical elements to interior-vertices and multiplicity to interior-edges to determine a chemical graph $G = (H, \alpha, \beta)$; and

- **C3.** Constraints for computing descriptors in the feature vector $f(G)$ of the selected chemical graph $G$.

In the constraints of C1, more formally we prepare the following.

**Variables:**

- a binary variable $v^X(i) \in \{0, 1\}$ for each vertex $v^X \in V_X$, $X \in \{C, T, F\}$ so that $v^X(i) = 1$ ⇔ vertex $v^X_i$ is used in a graph $H$ selected from $SG$;

- a binary variable $e^X(i) \in \{0, 1\}$ (resp., $e^C(i) \in \{0, 1\}$) for each edge $e^X_i \in E_T \cup E_F$ (resp., $e^C_i = a_i \in E_{(\geq 2)} \cup E_{(\geq 1)} \cup E_{(0/1)}$) so that $e^X(i) = 1$ ⇔ edge $e^X_i$ is used in a graph $H$ selected from $SG$. To save the number of variables in our MILP formulation, we do not prepare a binary variable $e^X(i, j) \in \{0, 1\}$ for any edge $e^X_{i, j} \in E_{CT} \cup E_{TF} \cup E_{CF} \cup E_{TC}$, where we represent a choice of edges in these sets by a set of $O(n^*|E_C|)$ variables (see Supplementary Materials for the details);

- binary variables $\delta^C_{i}(i, \psi) \in \{0, 1\}, i \in [1, t_C], \psi \in \mathcal{F}(v), v = v^C_i \in V_C$ and $\delta^F_{i}(i, \psi) \in \{0, 1\}, i \in [1, t_F], \psi \in \mathcal{F}_E$, where $\delta^X_{i}(i, \psi) = 1 \ (X \in \{C, T, F\})$ if and only if the $\rho$-fringe-tree rooted at vertex $v^X_i$ is $r$-isomorphic to $\psi$.

**Constraints:**

- linear constraints so that each $\rho$-fringe-tree rooted at a vertex $v^X_i$ in a graph $H$ from $SG$ is selected from the given set $\mathcal{F}(v^C_i)$ for $X = C$ (or $\mathcal{F}_E$ for $X \in \{T, F\}$);

- linear constraints such that each edge $e^C_i = a_i \in E_{(= 1)}$ is always used as an edge in $H$ and each edge $e^C_i = a_i \in E_{(0/1)}$ is used as an edge in $H$ if necessary;
- linear constraints such that for each edge \(a_k = (v^C_i, v^C_{i'}) \in E_{(\geq 2)}\), vertex \(v^C_i \in V_C\) is connected to vertex \(v^C_{i'} \in V_C\) in \(H\) by a pure path \(P_k\) that passes through some vertices in \(V_T\) and edges \(e^{CT}_{i,j}, e^T_{j+1,i}, e^T_{j+2,i}, \ldots, e^T_{j+p,i}, e^{TC}_{i',j+p}\) for some integers \(j\) and \(p\);

- linear constraints such that for each edge \(a_k = (v^C_i, v^C_{i'}) \in E_{(\geq 1)}\), either the edge \(a_k\) is used as an edge in \(H\) or vertex \(v^C_i \in V_C\) is connected to vertex \(v^C_{i'} \in V_C\) in \(H\) by a pure path \(P_k\) as in the case of edges in \(E_{(\geq 2)}\);

- linear constraints for selecting a leaf path \(Q_v\) rooted at a vertex \(v = v^C_i\) (resp., \(v = v^T_i\)) with \(\rho\)-internal edges \(e^{CF}_{i,j}\) (resp., \(e^{TF}_{i,j}\)), \(e^F_{j+1,i}, e^F_{j+2,i}, \ldots, e^F_{j+p,i}\) for some integers \(j\) and \(p\).

In the constraints of C2, we prepare an integer variable \(\alpha^X(i)\) for each vertex \(v^X_i \in V, X \in \{C, T, F\}\) in the scheme graph that represents the chemical element \(\alpha(v^X_i) \in \Lambda\) if \(v^X_i\) is in a selected graph \(H\) (or \(\alpha(v^X_i) = 0\) otherwise); integer variables \(\beta^C : E_C \rightarrow [0, 3], \beta^T : E_T \rightarrow [0, 3]\) and \(\beta^F : E_F \rightarrow [0, 3]\) that represent the bond-multiplicity of edges in \(E_C \cup E_T \cup E_F\); and integer variables \(\beta^+, \beta^- : E_{(\geq 2)} \cup E_{(\geq 1)} \rightarrow [0, 3]\) and \(\beta^{in} : V_C \cup V_T \rightarrow [0, 3]\) that represent the bond-multiplicity of edges in \(E_{CT} \cup E_{TC} \cup E_{CF} \cup E_{TF}\). This determines a chemical graph \(G = (H, \alpha, \beta)\). Also we include constraints for a selected chemical graph \(G\) to satisfy the valence condition at each interior-vertex \(v\) with the edge-configurations \(ec(e)\) of the edges \(e\) incident to \(v\) and the chemical specification \(\sigma_{ce}\).

In the constraints of C3, we introduce a variable for each descriptor and constraints with some more variables to compute the value of each descriptor in \(f(G)\) for a selected chemical graph \(G\).

The details of the MILP can be found in Section 3.

2 A Dynamic Programming Algorithm for Generating Isomers in Stage 5

![Figure 2: An illustration of a chemical graph \(G\), where for \(\rho = 2\), the exterior-vertices are \(w_1, w_2, \ldots, w_{19}\) and the interior-vertices are \(u_1, u_2, \ldots, u_{28}\).](image-url)

This section briefly reviews the method [4] for Stage 5. Let \(G^\dagger\) be a chemical graph that is a \((\sigma_{int}, \sigma_{ce})\)-extension of a seed graph \(G_C = (V_C, E_C)\), where we denote by \(E_{(=0)}\) the set of the edges
in $E_{(0/1)}$ that are not used in $G^\dagger$. We define a base-graph $G_B = (V_B, E_B)$ to be the seed graph $(V_C, E_C \setminus E_{(0)})$ after removing the edges in $E_{(=0)}$. We call a chemical graph $G^*$ a chemical isomer of $G^\dagger$ if $f(G^*) = f(G^\dagger)$ and $G^*$ is also a $(\sigma_{int}, \sigma_{ce})$-extension of $G_B$.

The method generates chemical isomers $G^*$ of $G^\dagger$ in the following way, where Figure 3 illustrates the whole process in the case of $V_B = \{v_1, v_2\}$ and $E_B = \{a_1, a_2\}$.

1. We first decompose a given chemical graph $G^\dagger$ into a collection of chemical rooted or bi-rooted trees.

   - For each vertex $v \in V_B$, let $T^\dagger_v$ denote the chemical rooted tree rooted at $v$ in $G$ that is constructed with a leaf path $Q_v$ and fringe-trees attached to $Q_v$. Possibly $T^\dagger_v$ consists of a single vertex $v$ and we call such a tree trivial.

   - For each edge $a = uv \in E_{(\geq 2)} \cup E_{(\geq 1)}$, let $T^a_{uv}$ denote the chemical bi-rooted tree rooted at vertices $u$ and $v$ in $G$ that consists of a pure $u,v$-path $P_a$, leaf paths rooted at internal vertices in $P_a$ and fringe-trees attached to these leaf paths. Possibly $T^a_v$ consists of a single edge $a$ and we call such a tree trivial.

Figure 4 illustrates the non-trivial chemical trees $T^\dagger_t, t \in V^*_B \cup E^*_B$ of the $(\sigma_{int}, \sigma_{ce})$-extension $G^\dagger = G$ in Figure 2.

2. Let $V^*_B$ (resp., $E^*_B$) denote the set of vertices $v \in V_B$ (resp., $a \in E_B$) such that $T^\dagger_v$ (resp., $T^a_{uv}$) is not trivial. For each vertex or edge $t \in V^*_B \cup E^*_B$, compute the feature vector $x_t^* = f(T^*_t)$ and then generate a set $T_t$ of all (or a limited number of) chemical acyclic graphs $T^*_t$ such that $f(T^*_t) = x_t^*$ and the structure of $T^*_t$ satisfies the lower and upper bounds in the interior-specification $\sigma_{int}$ by using the dynamic programming algorithm for chemical acyclic graphs [3].

3. For each combination of chemical trees $T^*_t, t \in V^*_B \cup E^*_B$, a chemical graph $G^*$ such that $f(G^*) = f(G^\dagger)$ is obtained from $G^\dagger$ by replacing each tree $T^\dagger_t$ with a new tree $T^*_t$. The number of such combinations is $\prod_{t \in V^*_B \cup E^*_B} |T_t|$, where we ignore a possible automorphism of the resulting graphs $G^*$.

The above method [4] can be used to generate chemical isomers in Stage 5 in our two-layered model by making a minor modification to the definition of a feature vector $f(G)$. 

![Figure 3: An illustration of generating a chemical isomer $G^*$ of a chemical graph $G^\dagger$ with a base-graph $G_B = (V_B, E_B)$.](image)

![Figure 4: Illustrates the non-trivial chemical trees $T^\dagger_t, t \in V^*_B \cup E^*_B$ of the $(\sigma_{int}, \sigma_{ce})$-extension $G^\dagger = G$ in Figure 2.](image)
Recall that

\[ T_v^\uparrow \text{ for } v \in \{u_5, u_{12}, u_{23}\} = V_B^* \]

and the non-trivial chemical bi-rooted trees \( T_a^\uparrow \) for \( a \in \{a_1 = u_1u_2, a_2 = u_1u_3, a_3 = u_4u_7, a_4 = u_{10}u_{11}, a_5 = u_{11}u_{12}\} = E_B^* \)

for the \((\sigma_{\text{int}}, \sigma_{\text{co}})\)-extension \( G^\uparrow = G \) in Figure 2, where the gray squares indicate the roots of these rooted and bi-rooted trees.

## 3 All Constraints in an MILP Formulation for Chemical Graphs

We define a standard encoding of a finite set \( A \) of elements to be a bijection \( \sigma : A \to [1, |A|] \), where we denote by \([A]\) the set \([1, |A|]\) of integers and by \([e]\) the encoded element \( \sigma(e) \). Let \( \epsilon \) denote \textit{null}, a fictitious chemical element that does not belong to any set of chemical elements, chemical symbols, adjacency-configurations and edge-configurations in the following formulation. Given a finite set \( A \), let \( A_\epsilon \) denote the set \( A \cup \{\epsilon\} \) and define a standard encoding of \( A_\epsilon \) to be a bijection \( \sigma : A \to [0, |A|] \) such that \( \sigma(\epsilon) = 0 \), where we denote by \([A]\) the set \([0, |A|]\) of integers and by \([e]\) the encoded element \( \sigma(e) \), where \([\epsilon]\) = 0.

### 3.1 Selecting a Cyclical-base

Recall that

\[
E_{(=1)} = \{ e \in E_C \mid \ell_{\text{LB}}(e) = \ell_{\text{UB}}(e) = 1 \}; \quad E_{(0/1)} = \{ e \in E_C \mid \ell_{\text{LB}}(e) = 0, \ell_{\text{UB}}(e) = 1 \};
\]

\[
E_{(\geq 1)} = \{ e \in E_C \mid \ell_{\text{LB}}(e) = 1, \ell_{\text{UB}}(e) \geq 2 \}; \quad E_{(\geq 2)} = \{ e \in E_C \mid \ell_{\text{LB}}(e) \geq 2 \};
\]

- Every edge \( a_i \in E_{(=1)} \) is included in \( G \);
- Each edge \( a_i \in E_{(0/1)} \) is included in \( G \) if necessary;
- For each edge \( a_i \in E_{(\geq 2)} \), edge \( a_i \) is not included in \( G \) and instead a path

\[
P_i = (v_{\text{tail}(i)}^C, v_{T_{j-1}}^T, v_{T_{j}}^T, \ldots, v_{T_{j+t}}^T, v_{\text{head}(i)}^C)
\]

of length at least 2 from vertex \( v_{\text{tail}(i)}^C \) to vertex \( v_{\text{head}(i)}^C \) visiting some vertices in \( V_T \) is constructed in \( G \); and
- For each edge \( a_i \in E_{(\geq 1)} \), either edge \( a_i \) is directly used in \( G \) or the above path \( P_i \) of length at least 2 is constructed in \( G \).
Let $t_C \triangleq |V_C|$ and denote $V_C$ by \{ $v^{C_i} \mid i \in [1, t_C]$\}. Regard the seed graph $G_C$ as a digraph such that each edge $a_i$ with end-vertices $v^{C_j}$ and $v^{C_j'}$ is directed from $v^{C_j}$ to $v^{C_j'}$ when $j < j'$. For each directed edge $a_i \in E_C$, let head$(i)$ and tail$(i)$ denote the head and tail of $e^C(i)$; i.e., $a_i = (v^{C_{\text{tail}(i)}}, v^{C_{\text{head}(i)}})$.

Assume that $E_C = \{ a_i \mid i \in [1, m_C] \}$, $E_{(2)} = \{ a_k \mid k \in [1, p] \}$, $E_{(1)} = \{ a_k \mid k \in [p + 1, q] \}$, $E_{(0/1)} = \{ a_i \mid i \in [q + 1, t] \}$ and $E_{(1)} = \{ a_i \mid i \in [t + 1, m_C] \}$ for integers $p, q$ and $t$. Let $I_{(1)}$ denote the set of indices $i$ of edges $a_i \in E_{(1)}$. Similarly for $I_{(0/1)}$, $I_{(1)}$ and $I_{(2)}$.

Define

$$k_C \triangleq \vert E_{(2)} \cup E_{(1)} \vert, \quad \bar{k}_C \triangleq \vert E_{(2)} \vert.$$

To control the construction of such a path $P_k$ for each edge $a_i \in E_{(2)} \cup E_{(1)}$, we regard the index $k \in [1, k_C]$ of each edge $a_k \in E_{(2)} \cup E_{(1)}$ as the “color” of the edge. To introduce necessary linear constraints that can construct such a path $P_k$ properly in our MILP, we assign the color $k$ to the vertices $v^{T}_{j-1}$, $v^{T}_j$, $v^{T}_{j+1}$ in $V_T$ when the above path $P_k$ is used in $G$.

For each index $s \in [1, t_C]$, let $I_C(s)$ denote the set of edges $e \in E_C$ incident to vertex $v^{C_s}$, and $E_{(1)}(s)$ (resp., $E_{(1)}(s)$) denote the set of edges $a_i \in E_{(1)}$ such that the tail (resp., head) of $a_i$ is vertex $v^{C_s}$. Similarly for $E_{(0/1)}(s)$, $E_{(0/1)}(s)$, $E_{(1)}(s)$, $E_{(1)}(s)$, $E_{(1)}(s)$ and $E_{(2)}(s)$. Let $I_C(s)$ denote the set of indices $i$ of edges $a_i \in I_C(s)$. Similarly for $I_{(1)}(s)$, $I_{(1)}(s)$, $I_{(1)}(s)$, $I_{(1)}(s)$, $I_{(2)}(s)$ and $I_{(2)}(s)$. Note that $[1, k_C] = I_{(2)} \cup I_{(1)}$ and $[k_C + 1, m_C] = I_{(1)} \cup I_{(0/1)} \cup I_{(1)}$.

**constants:**

- $t_C = |V_C|$, $k_C = |E_{(2)}|$, $k_C = |E_{(2)} \cup E_{(1)}|$, $t_T = n^\text{int}_{UB} - |V_C|$, $m_C = |E_C|$. Note that $a_i \in E_C \setminus (E_{(2)} \cup E_{(1)})$ holds $i \in [k_C + 1, m_C]$;

- $\ell_{\text{LB}}(k), \ell_{\text{UB}}(k) \in [1, t_T]$, $k \in [1, k_C]$: lower and upper bounds on the length of path $P_k$;

**variables:**

- $e^C(i) \in [0, 1], i \in [1, m_C]$: $e^C(i)$ represents edge $a_i \in E_C$, $i \in [1, m_C]$ ($e^C(i) = 1, i \in I_{(1)}$);

- $v^{T}(i) \in [0, 1], i \in [1, t_T]$: $v^{T}(i) = 1 \iff$ vertex $v^{T}_i$ is used in $G$;

- $e^{T}(i) \in \{0, 1\}, i \in [1, t_T+1]$: $e^{T}(i)$ represents edge $e^{T}_i = (v^{T}_{i-1}, v^{T}_i) \in E_T$, where $e^{T}_1$ and $e^{T}_{t_T+1}$ are fictitious edges ($e^{T}(i) = 1 \iff$ edge $e^{T}_i$ is used in $G$);

- $\chi^{T}(i) \in [0, k_C], i \in [1, t_T]$: $\chi^{T}(i)$ represents the color assigned to vertex $v^{T}_i$ ($\chi^{T}(i) = k > 0 \iff$ vertex $v^{T}_i$ is assigned color $k$; $\chi^{T}(i) = 0$ means that vertex $v^{T}_i$ is not used in $G$);

- $\text{clr}^{T}(k) \in [\ell_{\text{LB}}(k) - 1, \ell_{\text{UB}}(k) - 1]$, $k \in [1, k_C]$, $\text{clr}^{T}(0) \in [0, t_T]$: the number of vertices $v^{T}_i \in V_T$ with color $c$;

- $\delta^{T}_\chi(k) \in [0, 1], k \in [0, k_C]$: $\delta^{T}_\chi(k) = 1 \iff \chi^{T}(i) = k$ for some $i \in [1, t_T]$;

- $\chi^{T}(i, k) \in [0, 1], i \in [1, t_T], k \in [0, k_C]$ ($\chi^{T}(i, k) = 1 \iff \chi^{T}(i) = k$);

- $\text{deg}^{C}(i) \in [0, 4], i \in [1, t_C]$: the out-degree of vertex $v^{C}_i$ with the used edges $e^C$ in $E_C$;

- $\text{deg}^{C}(i) \in [0, 4], i \in [1, t_C]$: the in-degree of vertex $v^{C}_i$ with the used edges $e^C$ in $E_C$;
3.2 Constraints for Including Leaf Paths

Let \( \tilde{t}_C \) denote the number of vertices \( u \in V_C \) such that \( \text{bl}_{UB}(u) = 1 \) and assume that \( V_C = \{u_1, u_2, \ldots, u_p\} \) so that

\[
\text{bl}_{UB}(u_i) = 1, \quad i \in [1, \tilde{t}_C] \quad \text{and} \quad \text{bl}_{UB}(u_i) = 0, \quad i \in [\tilde{t}_C + 1, t_C].
\]

Define the set of colors for the vertex set \( \{u_i \mid i \in [1, \tilde{t}_C]\} \cup V_T \) to be \([1, c_F]\) with

\[
c_F \triangleq \tilde{t}_C + t_T = |\{u_i \mid i \in [1, \tilde{t}_C]\} \cup V_T|.
\]

Let each vertex \( v_C^i, i \in [1, \tilde{t}_C] \) (resp., \( v_T^i \in V_T \)) correspond to a color \( i \in [1, c_F] \) (resp., \( i + \tilde{t}_C \in [1, c_F] \)). When a path \( P = (u, v_F, v_F, \ldots, v_T) \) from a vertex \( u \in V_C \cup V_T \) is used in \( G \), we assign the color \( i \in [1, c_F] \) of the vertex \( u \) to the vertices \( v_F, v_F, \ldots, v_T \) in \( V_F \).

**constants:**

- \( c_F \): the maximum number of different colors assigned to the vertices in \( V_F \);
- \( n^\text{int}_{LB}, n^\text{int}_{UB} \in [2, n^*] \): lower and upper bounds on the number of interior-vertices in \( G \);
- \( \text{bl}_{LB}(i) \in [0, 1], i \in [1, \tilde{t}_C] \): a lower bound on the number of leaf \( \rho \)-branches in the leaf path rooted at a vertex \( v_C^i \);
- \( \text{bl}_{LB}(k), \text{bl}_{UB}(k) \in [0, \ell_{UB}(k) - 1], k \in [1, k_C] = I_{(\geq 2)} \cup I_{(\geq 1)} \): lower and upper bounds on the number of leaf \( \rho \)-branches in the trees rooted at internal vertices of a pure path \( P_k \) for an edge \( a_k \in E_{(\geq 1)} \cup E_{(\geq 2)} \);

**variables:**
- \( n^\text{int}_G \in [n^\text{int}_{LB}, n^\text{int}_{UB}] \): the number of interior-vertices in \( G \);
- \( v^F(i) \in [0, 1], i \in [1, t_F] \): \( v^F(i) = 1 \Leftrightarrow \) vertex \( v^F_i \) is used in \( G \);
- \( e^F(i) \in [0, 1], i \in [1, t_F + 1] \): \( e^F(i) \) represents edge \( e^F_i = v^F_{i-1}v^F_i \), where \( e^F_1 \) and \( e^F_{t_F+1} \) are fictitious edges (\( e^F(i) = 1 \Leftrightarrow \) edge \( e^F_i \) is used in \( G \));
- \( \chi^F(i) \in [0, c_F], i \in [1, t_F] \): \( \chi^F(i) \) represents the color assigned to vertex \( v^F_i \) (\( \chi^F(i) = c \Leftrightarrow \) vertex \( v^F_i \) is assigned color \( c \));
- \( c^\text{hF}(c) \in [0, t_F], c \in [0, c_F] \): the number of vertices \( v^F_i \) with color \( c \);
- \( \delta^\chi(c) \in [L^\text{LB}(c), 1], c \in [1, t_C] \): \( \delta^\chi(c) = 1 \Leftrightarrow \chi^F(i) = c \) for some \( i \in [1, t_F] \);
- \( \delta^\chi(c) \in [0, 1], c \in [t_C + 1, c_F] \): \( \delta^\chi(c) = 1 \Leftrightarrow \chi^F(i) = c \) for some \( i \in [1, t_F] \);
- \( \chi^F(i, c) \in [0, 1], i \in [1, t_F], c \in [0, c_F] \): \( \chi^F(i, c) = 1 \Leftrightarrow \chi^F(i) = c \);
- \( \text{bl}(k, i) \in [0, 1], k \in [1, k_C] = I_{(\geq 2)} \cup I_{(\geq 1)}, i \in [1, t_T] \): \( \text{bl}(k, i) = 1 \Leftrightarrow \) path \( P_k \) contains vertex \( v^T_i \) as an internal vertex and the \( p \)-fringe-tree rooted at \( v^T_i \) contains a leaf \( p \)-branch.

Constraints:

\[
\chi^F(i, 0) = 1 - v^F(i), \quad \sum_{c \in [0, c_F]} \chi^F(i, c) = 1, \quad \sum_{c \in [0, c_F]} c \cdot \chi^F(i, c) = \chi^F(i), \quad i \in [1, t_F], \quad (8)
\]

\[
\sum_{i \in [1, t_F]} \chi^F(i, c) = c^\text{hF}(c), \quad t_F \cdot \delta^\chi(c) \geq \sum_{i \in [1, t_F]} \chi^F(i, c) \geq \delta^\chi(c), \quad c \in [0, c_F], \quad (9)
\]

\[
e^F(1) = e^F(t_F + 1) = 0, \quad (10)
\]

\[
v^F(i - 1) \geq v^F(i),
\]

\[
c_F \cdot (v^F(i - 1) - e^F(i)) \geq \chi^F(i - 1) - \chi^F(i) \geq v^F(i - 1) - e^F(i), \quad i \in [2, t_F], \quad (11)
\]

\[
\text{bl}(k, i) \geq \delta^\chi(t_C + i) + \chi^T(i, k) - 1, \quad k \in [1, k_C], i \in [1, t_T], \quad (12)
\]

\[
\sum_{k \in [1, k_C], i \in [1, t_T]} \text{bl}(k, i) \leq \sum_{i \in [1, t_T]} \delta^\chi(t_C + i), \quad (13)
\]

\[
\text{bl}_{LB}(k) \leq \sum_{i \in [1, t_T]} \text{bl}(k, i) \leq \text{bl}_{UB}(k), \quad k \in [1, k_C], \quad (14)
\]

\[
t_C + \sum_{i \in [1, t_T]} v^T(i) + \sum_{i \in [1, t_F]} v^F(i) = n^\text{int}_G. \quad (15)
\]
3.3 Constraints for Including Fringe-trees

To express the condition that the $\rho$-fringe-tree is chosen from a rooted tree $C_i$, $T_i$ or $F_i$, we introduce the following set of variables and constraints.

constants:
- $n_{LB,n^*}$: lower and upper bounds on $n(G)$, where $n_{LB}, n^* \geq n_{LB}$;
- $ch_{LB}(i), ch_{UB}(i) \in [0,n^*]$, $i \in [1,t_T]$: lower and upper bounds on $ht(T_i)$ of the tree $T_i$ rooted at a vertex $v^C_i$;
- $ch_{LB}(k), ch_{UB}(k) \in [0,n^*]$, $k \in [1,k_C] = I_{(\geq 2)} \cup I_{(\geq 1)}$: lower and upper bounds on the maximum height $ht(T)$ of the tree $T \in F(P_k)$ rooted at an internal vertex of a path $P_k$ for an edge $a_k \in E_{(\geq 1)} \cup E_{(\geq 2)}$;
- Let $F_A$ denote the set of chemical rooted trees $\psi = (\{v\}, \emptyset)$ with $ht(\psi) = 0$ and $\alpha(v) = a$ for each chemical element $a \in \Lambda$;
- Prepare a coding of the set $F(D_r)$ and let $[\psi]$ denote the coded integer of an element $\psi$ in $F(D_r)$;
- Sets $F(v) \subseteq F(D_r)$, $v \in V_C$ and $F_E \subseteq F(D_r)$ of chemical rooted trees $T$ with $ht(T) \in [1, \rho]$;
- Define $F^* := \bigcup_{v \in V_C} F(v) \cup F_E$, $F_i^C := F(v^C_i)$, $i \in [1,t_C]$, $F_i^T := F_E$, $i \in [1,t_T]$ and $F_i^F := F_E$, $i \in [1,t_F]$;
- $F_i^X[p], p \in [1, \rho], X \in \{C,T,F\}$: the set of chemical rooted trees $T \in F_i^X$ with $ht(T) = p$;
- $n([\psi]) \in [0,3\rho]$, $\psi \in F^*$: the number of non-root vertices in a chemical rooted tree $\psi$;
- $ht([\psi]) \in [0, \rho]$, $\psi \in F^*$: the height of a chemical rooted tree $\psi$;
- $deg_r([\psi]) \in [0,4]$, $\psi \in F^*$: the number of children of the root $r$ of a chemical rooted tree $\psi$;

variables:
- $n_G \in [n_{LB}, n^*]$: $n(G)$;
- $v^X(i) \in [0,1], i \in [1,t_X], X \in \{T,F\}$: $v^X(i) = 1 \iff$ vertex $v^X_i$ is used in $G$;
- $h^X(i) \in [0, \rho], i \in [1,t_X], X \in \{C,T,F\}$: the height of the $\rho$-fringe-tree rooted at vertex $v^X_i$ in $G$;
- $\delta^X_H(i,[\psi]) \in [0,1], i \in [1,t_X], \psi \in F_i^X \cup F_A, X \in \{T,F\}$: $\delta^X_H(i,[\psi]) = 1 \iff \psi$ is the $\rho$-fringe-tree at vertex $v^X_i$, where $\psi \in F_A$ means that the height of the $\rho$-fringe-tree is 0;
- $deg^X_{ex}(i) \in [0,3], i \in [1,t_X], X \in \{C,T,F\}$: the number of children of the root of the $\rho$-fringe-tree rooted at vertex $v^X_i$ in $G$;
- $\sigma(k,i) \in [0,1], k \in [1,k_C] = I_{(\geq 2)} \cup I_{(\geq 1)}, i \in [1,t_T]$: $\sigma(k,i) = 1 \iff$ the $\rho$-fringe-tree $T_v$ rooted at vertex $v = v^T_i$ with color $k$ has the largest height among such trees;

constraints:

\[
\sum_{\psi \in F_i^C \cup F_A} \delta^C_H(i,[\psi]) = 1, \quad \sum_{\psi \in F_i^C \cup F_A} deg_r([\psi]) \cdot \delta^C_H(i,[\psi]) = deg^C_{ex}(i), \quad i \in [1,t_C],
\]

\[
\sum_{\psi \in F_i^X \cup F_A} \delta^X_H(i,[\psi]) = v^X(i), \quad \sum_{\psi \in F_i^X \cup F_A} deg_r([\psi]) \cdot \delta^X_H(i,[\psi]) = deg^X_{ex}(i), \quad i \in [1,t_X], X \in \{T,F\}, \quad (16)
\]
\[ \sum_{\psi \in F^i} \delta^F_i([\psi]) \geq v^F(i) - e^F(i + 1), \quad i \in [1, t_F] \quad (e^F(t_F + 1) = 0), \quad (17) \]

\[ \sum_{\psi \in F^i_X} \text{ht}([\psi]) \cdot \delta^X_X(i, [\psi]) = h^X(i), \quad i \in [1, t_X], \quad X \in \{C, T, F\}, \quad (18) \]

\[ \sum_{\psi \in F^i_K} n([\psi]) \cdot \delta^X(i, [\psi]) + \sum_{i \in [1, t_X], X \in \{T, F\}} v^X(i) + t_C = n_G, \quad (19) \]

\[ h^C(i) \geq c_{\text{LB}}(i) - n^*\delta^F(i), \quad \text{chr}^F(i) + \rho \geq c_{\text{LB}}(i), \]
\[ h^C(i) \leq c_{\text{UB}}(i), \quad \text{chr}^F(i) + \rho \leq c_{\text{UB}}(i) + n^*(1 - \delta^F(i)), \quad i \in [1, t_C], \quad (20) \]

\[ c_{\text{LB}}(i) \leq h^C(i) \leq c_{\text{UB}}(i), \quad i \in [k_C + 1, t_C], \quad (21) \]

\[ h^T(i) \leq c_{\text{UB}}(k) + n^*(\delta^F_X(t_C + i) + 1 - \chi^T(i, k)), \]
\[ \text{chr}^F(t_C + i) + \rho \leq c_{\text{UB}}(k) + n^*(2 - \delta^F_X(t_C + i) - \chi^T(i, k)), \quad k \in [1, k_C], \quad i \in [1, t_T], \quad (22) \]

\[ \sum_{i \in [1, t_T]} \sigma(k, i) = \delta^T_X(k), \quad k \in [1, k_C], \quad (23) \]

\[ \chi^T(i, k) \geq \sigma(k, i), \]
\[ h^T(i) \geq c_{\text{LB}}(k) - n^*(\delta^F_X(t_C + i) + 1 - \sigma(k, i)), \]
\[ \text{chr}^F(t_C + i) + \rho \geq c_{\text{LB}}(k) - n^*(2 - \delta^F_X(t_C + i) - \sigma(k, i)), \quad k \in [1, k_C], \quad i \in [1, t_T]. \quad (24) \]

3.4 Descriptor for the Number of Specified Degree

We include constraints to compute descriptors \( \text{deg}^m_{\text{all}}(G), \quad d \in [1, 4]. \)

**variables:**

- \( \text{deg}^X(i) \in [0, 4], \quad i \in [1, t_X], \quad X \in \{C, T, F\} \): the degree \( \text{deg}_G(v^X_i) \) of vertex \( v^X_i \) in \( G \);
- \( \text{deg}_{CT}(i) \in [0, 4], \quad i \in [1, t_C] \): the number of edges from vertex \( v^C_i \) to vertices \( v^T_j, \quad j \in [1, t_T] \);
- \( \text{deg}_{TC}(i) \in [0, 4], \quad i \in [1, t_C] \): the number of edges from vertices \( v^T_j, \quad j \in [1, t_T] \) to vertex \( v^C_i \);
Supplementary Materials

- \( \delta_{\text{dg}}^{X}(i, d) \in [0, 1], i \in [1, t_{C}], d \in [1, 4], \delta_{\text{dg}}^{X}(i, d) \in [0, 1], i \in [1, t_{X}], d \in [0, 4], X \in \{T, F\}: \delta_{\text{dg}}^{X}(i, d) = 1 \iff \deg_{\text{X}}^{G}(i) = d; \)
- \( \text{dg}(d) \in [\text{dg}_{\text{LB}}(d), \text{dg}_{\text{UB}}(d)], d \in [1, 4]: \) the number of interior-vertices \( v \) with \( \deg_{G}(v) = d; \)
- \( \text{deg}_{\text{int}}^{G}(i) \in [1, 4], i \in [1, t_{C}], \text{deg}_{\text{int}}^{G}(i) \in [0, 4], i \in [1, t_{X}], X \in \{T, F\}: \) the interior-degree \( \deg_{\{\text{int}, \text{ext}\}}^{G}(v_{i}); \) i.e., the number of interior-edges incident to vertex \( v_{X}^{i}; \)
- \( \delta_{\text{int}}^{G}(i, d) \in [0, 1], i \in [1, t_{C}], d \in [1, 4], \delta_{\text{int}}^{G}(i, d) \in [0, 1], i \in [1, t_{X}], d \in [0, 4], X \in \{T, F\}: \delta_{\text{int}}^{G}(i, d) = 1 \iff \deg_{\text{int}}^{G}(i) = d; \)
- \( \text{deg}_{\text{int}}^{G}(d) \in [\text{deg}_{\text{LB}}(d), \text{deg}_{\text{UB}}(d)], d \in [1, 4]: \) the number of interior-vertices \( v \) with the interior-degree \( \deg_{\{\text{int}, \text{ext}\}}^{G}(v) = d; \)

**constraints:**

\[
\sum_{k \in I_{(\geq 2)}^{+}(i) \cup J_{(\geq 1)}^{+}(i)} \delta_{k}^{T} = \deg_{\text{CT}}^{G}(i), \quad \sum_{k \in I_{(\geq 2)}^{+}(i) \cup J_{(\geq 1)}^{+}(i)} \delta_{k}^{T} = \deg_{\text{TC}}^{G}(i), \quad i \in [1, t_{C}],
\]

\[
\delta_{\text{int}}^{C}(i) = \frac{v^{T}(i) + \delta_{F}(\bar{t}_{C} + i)}{2}, \quad \deg_{\text{int}}^{G}(i) = \frac{v^{F}(i) + e^{T}(1)}{2}, \quad i \in [1, t_{C}],
\]

\[
\delta_{\text{int}}^{C}(i) = \frac{v^{F}(i) + e^{T}(i + 1)}{2}, \quad \deg_{\text{int}}^{G}(i) = \frac{v^{T}(i) + \delta_{F}(\bar{t}_{C} + i)}{2}, \quad i \in [1, t_{C}],
\]

\[
\sum_{d \in [0, 4]} \delta_{\text{int}}^{C}(i, d) = 1, \quad \sum_{d \in [1, 4]} d \cdot \delta_{\text{int}}^{C}(i, d) = \deg_{\text{C}}^{G}(i),
\]

\[
\sum_{d \in [0, 4]} \delta_{\text{int}}^{G}(i, d) = 1, \quad \sum_{d \in [1, 4]} d \cdot \delta_{\text{int}}^{G}(i, d) = \deg_{\text{X}}^{G}(i), \quad i \in [1, t_{X}], X \in \{T, C, F\},
\]

\[
\sum_{i \in [1, t_{C}]} \delta_{\text{int}}^{C}(i, d) + \sum_{i \in [1, t_{T}]} \delta_{\text{int}}^{T}(i, d) + \sum_{i \in [1, t_{F}]} \delta_{\text{int}}^{F}(i, d) = \text{dg}(d), \quad d \in [1, 4].
\]
3.5 Assigning Multiplicity

We prepare an integer variable $\beta(e)$ for each edge $e$ in the scheme graph $SG$ to denote the bond-multiplicity of $e$ in a selected graph $G$ and include necessary constraints for the variables to satisfy in $G$.

**constants:**

- $\beta_i([\psi])$: the sum of bond-multiplicities of edges incident to the root of a tree $\psi \in F^*$;

**variables:**

- $\beta^X(i) \in [0, 3], i \in [2, t_X], X \in \{T, F\}$: the bond-multiplicity of edge $e^X_i$;
- $\beta^C(i) \in [0, 3], i \in [\overline{k}_C + 1, m_C] = I_{(\geq 1)} \cup I_{(0/1)} \cup I_{(=1)}$: the bond-multiplicity of edge $a_i \in E_{(\geq 1)} \cup E_{(0/1)} \cup E_{(=1)}$;
- $\beta^+(k), \beta^-(k) \in [0, 3], k \in [1, k_C] = I_{(\geq 2)} \cup I_{(\geq 1)}$: the bond-multiplicity of the first (resp., last) edge of the pure path $P_k$;
- $\beta^{\text{in}}(c) \in [0, 3], c \in [1, c_F]$: the bond-multiplicity of the first edge of the leaf path $Q_c$ rooted at vertex $c$;
- $\beta^{\text{ex}}(i) \in [0, 4], i \in [1, t_X], X \in \{C, T, F\}$: the sum $\beta_{T^i}(v)$ of bond-multiplicities of edges in the $\rho$-fringe-tree $T_v$ rooted at interior-vertex $v = v^X_i$;
- $\delta^{X}(i, m) \in [0, 1], i \in [2, t_X], m \in [0, 3], X \in \{T, F\}$: $\delta^{X}(i, m) = 1 \Leftrightarrow \beta^{X}(i) = m$;
- $\delta^{C}(i, m) \in [0, 1], i \in [\overline{k}_C, m_C] = I_{(\geq 1)} \cup I_{(0/1)} \cup I_{(=1)}, m \in [0, 3], X \in \{T, F\}$: $\delta^{C}(i, m) = 1 \Leftrightarrow \beta^{C}(i) = m$;
- $\delta^{+}(k), \delta^{-}(k) \in [0, 1], k \in [1, k_C] = I_{(\geq 2)} \cup I_{(\geq 1)}, m \in [0, 3]$: $\delta^{+}(k, m) = 1 \Leftrightarrow \delta^{-}(k, m) = 1$ (resp., $\delta^{+}(k, m) = 1 \Leftrightarrow \delta^{-}(k, m) = 1$);
- $\delta^{\text{in}}(c, m) \in [0, 1], c \in [1, c_F], m \in [0, 3]$: $\delta^{\text{in}}(c, m) = 1 \Leftrightarrow \beta^{\text{in}}(c) = m$;
- $\text{bd}^{\text{int}}(m) \in [0, 2n_{\text{UB}}^{\text{int}}], m \in [1, 3]$: the number of interior-edges with bond-multiplicity $m$ in $G$;
- $\text{bd}^X(m) \in [0, 2n_{\text{UB}}^{\text{int}}], m \in [0, 2n_{\text{UB}}^{\text{int}}], X \in \{C, T, CT, TC\}$, $\text{bd}^X(m) \in [0, 2n_{\text{UB}}^{\text{int}}], m \in [1, 3]$: the number of interior-edges $e \in E_X$ with bond-multiplicity $m$ in $G$;

**constraints:**

\[ e^C(i) \leq \beta^C(i) \leq 3e^C(i), i \in [\overline{k}_C + 1, m_C] = I_{(\geq 1)} \cup I_{(0/1)} \cup I_{(=1)}, \]  
(34)

\[ e^X(i) \leq \beta^X(i) \leq 3e^X(i), \quad i \in [2, t_X], X \in \{T, F\}, \]  
(35)

\[ \delta^T_X(k) \leq \beta^+(k) \leq 3\delta^T_X(k), \quad \delta^T_X(k) \leq \beta^-(k) \leq 3\delta^T_X(k), \quad k \in [1, k_C], \]  
(36)

\[ \delta^F_X(c) \leq \beta^{\text{in}}(c) \leq 3\delta^F_X(c), \quad c \in [1, c_F], \]  
(37)
\[\sum_{m \in [0,3]} \delta_X^C(i,m) = 1, \quad \sum_{m \in [0,3]} m \cdot \delta_X^C(i,m) = \beta_X(i), \quad i \in [2, t_X], X \in \{T, F\}, \quad (38)\]

\[\sum_{m \in [0,3]} \delta_Y^C(i,m) = 1, \quad \sum_{m \in [0,3]} m \cdot \delta_Y^C(i,m) = \beta_Y(i), \quad i \in [\tilde{k}_C + 1, m_C], \quad (39)\]

\[\sum_{m \in [0,3]} \delta_{+}^C(k,m) = 1, \quad \sum_{m \in [0,3]} m \cdot \delta_{+}^C(k,m) = \beta_{+}(k), \quad k \in [1, k_C],\]

\[\sum_{m \in [0,3]} \delta_{-}^C(k,m) = 1, \quad \sum_{m \in [0,3]} m \cdot \delta_{-}^C(k,m) = \beta_{-}(k), \quad k \in [1, k_C],\]

\[\sum_{m \in [0,3]} \delta_{-}^C(c,m) = 1, \quad \sum_{m \in [0,3]} m \cdot \delta_{-}^C(c,m) = \beta_{-}(c), \quad c \in [1, c_F], \quad (40)\]

\[\sum_{\psi \in F_X^i} \beta_i([\psi]) \cdot \delta_X^C(i, [\psi]) = \beta_X(i), \quad i \in [1, t_X], X \in \{C, T, F\}, \quad (41)\]

\[\sum_{i \in [\tilde{k}_C + 1, m_C]} \delta_X^C(i,m) = bd_C(m), \quad \sum_{i \in [2, t_T]} \delta_Y^C(i,m) = bd_T(m),\]

\[\sum_{k \in [1, k_C]} \delta_{+}^C(k,m) = bd_{CT}(m), \quad \sum_{k \in [1, k_C]} \delta_{-}^C(k,m) = bd_{TC}(m),\]

\[\sum_{i \in [2, t_F]} \delta_X^C(i,m) = bd_F(m), \quad \sum_{c \in [1, c_C]} \delta_{-}^C(c,m) = bd_{CF}(m),\]

\[\sum_{c \in [c_C + 1, c_F]} \delta_{-}^C(c,m) = bd_{TF}(m),\]

\[bd_C(m) + bd_T(m) + bd_F(m) + bd_{CT}(m) + bd_{TC}(m) + bd_{TF}(m) + bd_{CF}(m) = bd_{int}(m), \quad m \in [1, 3]. \quad (42)\]

### 3.6 Assigning Chemical Elements and Valence Condition

We include constraints so that each vertex \(u\) in a selected graph \(H\) satisfies the valence condition; i.e., \(\sum_{uv \in E(H)} \beta(uv) \leq \text{val}(\alpha(u))\). With these constraints, a chemical graph \(G = (H, \alpha, \beta)\) on a selected subgraph \(H\) will be constructed.

**Constants:**

- Subsets \(\Lambda^\text{int}, \Lambda^\text{ex} \subseteq \Lambda\) of chemical elements, where we denote by \([e]\) (resp., \([e]^\text{int}\) and \([e]^\text{ex}\)) of a standard encoding of an element \(e\) in the set \(\Lambda\) (resp., \(\Lambda^\text{int}\) and \(\Lambda^\text{ex}\));

- A valence function: \(\text{val} : \Lambda \rightarrow [1, 4]\);

- A function \(\text{mass}^* : \Lambda \rightarrow \mathbb{Z}\) (we let \(\text{mass}(a)\) denote the observed mass of a chemical element \(a \in \Lambda\), and define \(\text{mass}^*(a) \triangleq \lfloor 10 \cdot \text{mass}(a) \rfloor\)).
- Subsets $\Lambda^*(i) \subseteq \Lambda^\text{int}$, $i \in [1, t_C]$;
- $n_{\text{LB}}(a), n_{\text{UB}}(a) \in [0, n^*]$, $a \in \Lambda$: lower and upper bounds on the number of vertices $v$ with $\alpha(v) = a$;
- $n_{\text{LB}}^\text{int}(a), n_{\text{UB}}^\text{int}(a) \in [0, n^*]$, $a \in \Lambda^\text{int}$: lower and upper bounds on the number of interior-vertices $v$ with $\alpha(v) = a$;
- $\alpha_r([\psi]) \in [\Lambda^{\text{ex}},] \in \mathcal{F}^* \cup \mathcal{F}_\Lambda$: the chemical element $\alpha(r)$ of the root $r$ of $\psi$;
- $n_{\text{ex}}(\psi), \psi \in \mathcal{F}^*$: the frequency of chemical element $a$ in the set of non-rooted vertices in $\psi$;
- $n_{\text{h}}([\psi], d) \in [0, 3^t], \psi \in \mathcal{F}^* \cup \mathcal{F}_\Lambda, d \in [0, 3]$: the number of non-root vertices with $\deg_{\text{hyd}}(v) = d$ in $\psi$.

variables:
- $\beta^\text{CT}(i), \beta^\text{TC}(i) \in [0, 3], i \in [1, t_T]$: the bond-multiplicity of edge $e^{\text{CT}}_{j,i}$ (resp., $e^{\text{TC}}_{j,i}$) if one exists;
- $\beta^\text{CF}(i), \beta^\text{TF}(i) \in [0, 3], i \in [1, t_F]$: the bond-multiplicity of $e^{\text{CF}}_{j,i}$ (resp., $e^{\text{TF}}_{j,i}$) if one exists;
- $\alpha^X(i) \in [\Lambda^\text{int}], \delta^X(i, [a]^\text{int}) \in [0, 1], a \in \Lambda^\text{int}, i \in [1, t_X], X \in \{C, T, F\}$: $\alpha^X(i) = [a]^\text{int} \geq 1$ (resp., $\delta^X(i) = 1$ (resp., $\delta^X(i, 0) = 0$) $\iff \alpha(v^X_i) = a \in \Lambda$ (resp., vertex $v^X_i$ is not used in $G$);
- $\delta^X(i, [a]^\text{int}) \in [0, 1], i \in [1, t_X], a \in \Lambda^\text{int}, X \in \{C, T, F\}$: $\delta^X(i, [a]^t) = 1 \iff \alpha(v^X_i) = a$;
- Mass $\in \mathbb{Z}_+$: $\sum_{v \in V(H)} \text{mass}(\alpha(v))$;
- $n_{\text{a}}([a]) \in [n_{\text{LB}}(a), n_{\text{UB}}(a)], a \in \Lambda$: the number of vertices $v \in V(H)$ with $\alpha(v) = a$;
- $n_{\text{int}}([a]^\text{int}) \in [n_{\text{LB}}^\text{int}(a), n_{\text{UB}}^\text{int}(a)], a \in \Lambda, X \in \{C, T, F\}$: the number of interior-vertices $v \in V(G)$ with $\alpha(v) = a$;
- $n_{\text{ex}}^X([a]^\text{ex}), n_{\text{ex}}^X([a]^\text{ex}) \in [0, n_{\text{UB}}(a)], a \in \Lambda, X \in \{C, T, F\}$: the number of exterior-vertices rooted at vertices $v \in V_X$ and the number of exterior-vertices $v$ such that $\alpha(v) = a$;
- $\delta_{\text{hyd}}^X(i, d) \in [0, 1], d \in [0, 3], X \in \{C, T, F\}$: $\delta_{\text{hyd}}^X(i, d) \iff \deg_{\text{hyd}}(v^X_i) = d$;
- hyd($d), d \in [0, 3]$: the number of vertices $v$ with $\deg_{\text{hyd}}(v^X_i) = d$;

constraints:

$$
\begin{align*}
\beta^+(k) - 3(e^T(i) - \chi^T(i, k) + 1) \leq & \beta^\text{CT}(i) \leq \beta^+(k) + 3(e^T(i) - \chi^T(i, k) + 1), i \in [1, t_T], \\
\beta^-(k) - 3(e^T(i + 1) - \chi^T(i, k) + 1) \leq & \beta^\text{TC}(i) \leq \beta^-(k) + 3(e^T(i + 1) - \chi^T(i, k) + 1), i \in [1, t_T], \\
& k \in [1, k_C], \\
\beta^\text{in}(c) - 3(e^F(i) - \chi^F(i, c) + 1) \leq & \beta^\text{CF}(i) \leq \beta^\text{in}(c) + 3(e^F(i) - \chi^F(i, c) + 1), i \in [1, t_F], \\
\beta^\text{in}(c) - 3(e^F(i) - \chi^F(i, c) + 1) \leq & \beta^\text{TF}(i) \leq \beta^\text{in}(c) + 3(e^F(i) - \chi^F(i, c) + 1), i \in [1, t_F], \\
& c \in [1, t_C], \\
\beta^\text{in}(c) - 3(e^F(i) - \chi^F(i, c) + 1) \leq & \beta^\text{FC}(i) \leq \beta^\text{in}(c) + 3(e^F(i) - \chi^F(i, c) + 1), i \in [1, t_F], \\
& c \in [t_C + 1, t_F],
\end{align*}
$$

(43)

(44)
\[ \sum_{a \in A^{\text{int}}} \delta_a^X(i, [a]^{\text{int}}) = \begin{cases} 1, & \text{if } \alpha(i) = a \in A^{\text{int}}, \\ \sum_{a \in A^{\text{int}}} [a]^{\text{int}} \cdot \delta_a^X(i, [a]^{\text{int}}) = \alpha(i), & \text{if } i \in [1, t_C], \end{cases} \]

\[ \sum_{a \in A^{\text{int}}} \delta_a^X(i, [a]^{\text{int}}) = v_X(i), \quad \sum_{a \in A^{\text{int}}} [a]^{\text{int}} \cdot \delta_a^X(i, [a]^{\text{int}}) = \alpha^X(i), \quad i \in [1, t_C], X \in \{T, F\}, \quad (45) \]

\[ \sum_{\psi \in P_X \cup J_A} \alpha_r([\psi]) \cdot \delta_n^X(i, [\psi]) = \alpha^X(i), \quad i \in [1, t_C], X \in \{C, T, F\}, \quad (46) \]

\[ \sum_{j \in I_C(i)} \beta^C(j) + \sum_{k \in I_{(2/1)(i)}^{+} \cup I_{(2/1)(i)}^{-}} \beta^+(k) + \sum_{k \in I_{(-1/2)(i)}^{+} \cup I_{(-1/2)(i)}^{-}} \beta^-(k) \]
\[ + \beta_{\text{in}}(i) + \beta_{\text{ex}}^C(i) + \sum_{d \in [0, 3]} d \cdot \delta_{\text{hyd}}^C(i, d) = \sum_{a \in A^{\text{int}}} \text{val}(a) \delta_a^C(i, [a]^{\text{int}}), \quad i \in [1, t_C], \quad (47) \]

\[ \sum_{j \in I_C(i)} \beta^C(j) + \sum_{k \in I_{(2/1)(i)}^{+} \cup I_{(2/1)(i)}^{-}} \beta^+(k) + \sum_{k \in I_{(-1/2)(i)}^{+} \cup I_{(-1/2)(i)}^{-}} \beta^-(k) \]
\[ + \beta_{\text{ex}}^C(i) + \sum_{d \in [0, 3]} d \cdot \delta_{\text{hyd}}^C(i, d) = \sum_{a \in A^{\text{int}}} \text{val}(a) \delta_a^C(i, [a]^{\text{int}}), \quad i \in [t_C + 1, t_C], \quad (48) \]

\[ \beta^T(i) + \beta^T(i + 1) + \beta_{\text{ex}}^T(i) + \beta^{CT}(i) + \beta^{TC}(i) \]
\[ + \beta_{\text{in}}(t_C + i) + \sum_{d \in [0, 3]} d \cdot \delta_{\text{hyd}}^T(i, d) = \sum_{a \in A^{\text{int}}} \text{val}(a) \delta_a^T(i, [a]^{\text{int}}), \]
\[ i \in [1, t_T] \quad (\beta^T(1) = \beta^T(t_T + 1) = 0), \quad (49) \]

\[ \beta^F(i) + \beta^F(i + 1) + \beta_{\text{ex}}^F(i) + \beta^{CF}(i) + \beta^{TF}(i) \]
\[ + \beta_{\text{ex}}^F(i) + \sum_{d \in [0, 3]} d \cdot \delta_{\text{hyd}}^F(i, d) = \sum_{a \in A^{\text{int}}} \text{val}(a) \delta_a^F(i, [a]^{\text{int}}), \]
\[ i \in [1, t_F] \quad (\beta^F(1) = \beta^F(t_F + 1) = 0), \quad (50) \]

\[ \sum_{i \in [1, t_X]} \delta_a^X(i, [a]^{\text{int}}) = n_a^X([a]^{\text{int}}), \quad a \in A^{\text{int}}, X \in \{C, T, F\}, \quad (51) \]

\[ \sum_{\psi \in P_X} n_a^\text{ex}_{\alpha}(i, [\psi]) \cdot \delta_n^X(i, [\psi]) = n_a^\text{ex}_{X}([a]^{\text{ex}}), \quad a \in A^{\text{ex}}, X \in \{C, T, F\}, \quad (52) \]
naC([a]^{\text{int}}) + naT([a]^{\text{int}}) + naF([a]^{\text{int}}) = na^{\text{int}}([a]), \quad a \in \Lambda^{\text{int}},
\sum_{X \in \{C,T,F\}} na_c^{\text{ex}}([a]^{\text{ex}}) = na^{\text{ex}}([a]), \quad a \in \Lambda^{\text{ex}},
na^{\text{int}}([a]) + na^{\text{ex}}([a]) = na([a]), \quad a \in \Lambda^{\text{int}} \cap \Lambda^{\text{ex}},
na^{\text{int}}([a]) = na([a]), \quad a \in \Lambda^{\text{int}} \setminus \Lambda^{\text{ex}},
na^{\text{ex}}([a]) = na([a]), \quad a \in \Lambda^{\text{ex}} \setminus \Lambda^{\text{int}}, \quad (53)
\sum_{a \in \Lambda} \text{mass}^*(a) \cdot na([a]) = \text{Mass}, \quad (54)
\sum_{d \in [0,3]} \delta_c^{\text{hyd}}(i,d) = 1, i \in [1,tC],
\sum_{d \in [0,3]} \delta_X^{\text{hyd}}(i,d) = v_X(i), i \in [1,tX], X \in \{T,F\}, \quad (55)
\sum_{i \in [1,tX], X \in \{C,T,F\}} \delta^{\text{hyd}}(i,d) + \sum_{\psi \in P_X, i \in [1,tX], X \in \{C,T,F\}} n_{\mathbb{H}}([\psi], d) \cdot \delta^{X}(i,[\psi]) = \text{hydg}(d), d \in [0,3], \quad (56)
\sum_{a \in \Lambda^*(i)} \delta^a(i,[a]) = 1, \quad i \in [1,tC]. \quad (57)

3.7 Constraints for Bounds on the Number of Bonds

We include constraints for specification of lower and upper bounds bd_{LB} and bd_{UB}.

constants:
- bd_{m,LB}(i), bd_{m,UB}(i) \in [0,n_{\text{int}}], i \in [1,m_C], m \in [2,3]: lower and upper bounds on the number of edges e \in E(P_i) with bond-multiplicity \beta(e) = m in the pure path P_i for edge e_i \in E_C;

variables :
- bd_T(k,i,m) \in [0,1], k \in [1,k_C], i \in [2,t_T], m \in [2,3]: bd_T(k,i,m) = 1 \iff the pure path P_k for edge e_k \in E_C contains edge e^T_i with \beta(e^T_i) = m;

constraints:
bd_{m,LB}(i) \leq \delta^C_{\beta}(i,m) \leq bd_{m,UB}(i), i \in I_{(=1)} \cup I_{(0/1)}, m \in [2,3], \quad (58)
bd_T(k,i,m) \geq \delta^T_{\beta}(i,m) + \chi^T(i,k) - 1, \quad k \in [1,k_C], i \in [2,t_T], m \in [2,3], \quad (59)
\[
\sum_{j \in [2, t_T]} \delta^j_B(j, m) \geq \sum_{k \in [1, k_C], j \in [2, t_T]} \text{bd}_T(k, i, m), \quad m \in [2, 3], \quad (60)
\]
\[
\text{bd}_{m, \text{LB}}(k) \leq \sum_{i \in [2, t_T]} \text{bd}_T(k, i, m) + \delta^+_B(k, m) + \delta^-_B(k, m) \leq \text{bd}_{m, \text{UB}}(k),
\]
\[
k \in [1, k_C], m \in [2, 3]. \quad (61)
\]

### 3.8 Descriptor for the Number of Adjacency-configurations

We call a tuple \((a, b, m) \in \Lambda \times \Lambda \times [1, 3]\) an adjacency-configuration. The adjacency-configuration of an edge-configuration \((\mu = ad, \mu' = bd', m)\) is defined to be \((a, b, m)\). We include constraints to compute the frequency of each adjacency-configuration in an inferred chemical graph \(G\).

**constants:**

- A set \(\Gamma^\text{int}\) of edge-configurations \(\gamma = (\mu, \xi, m)\) with \(m \leq \xi\);
- Let \(\overline{\gamma}\) of an edge-configuration \(\gamma = (\mu, \xi, m)\) denote the edge-configuration \((\xi, \mu, m)\);
- Let \(\Gamma^\text{int}_{<} = \{(\mu, \xi, m) \in \Gamma^\text{int} \mid \mu < \xi\}\), \(\Gamma^\text{int}_{=} = \{(\mu, \xi, m) \in \Gamma^\text{int} \mid \mu = \xi\}\) and \(\Gamma^\text{int}_{>} = \{\gamma \mid \gamma \in \Gamma^\text{int}_{<}\}\);
- Let \(\Gamma^\text{int}_{ac,<}, \Gamma^\text{int}_{ac,=}, \text{ and } \Gamma^\text{int}_{ac,>}\) denote the sets of the adjacency-configurations of edge-configurations in the sets \(\Gamma^\text{int}_{<}, \Gamma^\text{int}_{=}\) and \(\Gamma^\text{int}_{=},\) respectively;
- Let \(\bar{\nu}\) of an adjacency-configuration \(\nu = (a, b, m)\) denote the adjacency-configuration \((b, a, m)\);
- Prepare a coding of the set \(\Gamma^\text{int}_{ac} \cup \Gamma^\text{int}_{ac,=}\) and let \([\nu]^\text{int}\) denote the coded integer of an element \(\nu\) in \(\Gamma^\text{int}_{ac} \cup \Gamma^\text{int}_{ac,=}\);
- Choose subsets \(\Gamma^\text{FC}_{ac}, \Gamma^\text{CT}_{ac}, \Gamma^\text{TC}_{ac}, \Gamma^\text{FC}_{ac}, \Gamma^\text{CF}_{ac}, \Gamma^\text{FF}_{ac} \subseteq \Gamma^\text{int}_{ac} \cup \Gamma^\text{int}_{ac,=}\). To compute the frequency of adjacency-configurations exactly, set \(\tilde{\Gamma}^\text{FC}_{ac} := \tilde{\Gamma}^\text{CT}_{ac} := \tilde{\Gamma}^\text{TC}_{ac} := \tilde{\Gamma}^\text{FC}_{ac} := \tilde{\Gamma}^\text{CF}_{ac} := \tilde{\Gamma}^\text{FF}_{ac} := \Gamma^\text{int}_{ac} \cup \Gamma^\text{int}_{ac,=}\);
- \(ac^\text{int}_{\text{LB}}(\nu), ac^\text{int}_{\text{UB}}(\nu) \in [0, 2n^\text{int}_{\text{UB}}], \nu = (a, b, m) \in \Gamma^\text{int}_{ac},\) lower and upper bounds on the number of interior-edges \(e = uv\) with \(\alpha(u) = a, \alpha(v) = b\) and \(\beta(e) = m\);

**variables:**

- \(ac^\text{int}(\nu)^\text{int} \in [ac^\text{int}_{\text{LB}}(\nu), ac^\text{int}_{\text{UB}}(\nu)], \nu \in \Gamma^\text{int}_{ac}:\) the number of interior-edges with adjacency-configuration \(\nu\);
- \(ac^\text{CT}(\nu)^\text{int} \in [0, m^\text{CT}], \nu \in \tilde{\Gamma}^\text{CT}_{ac}, ac^\text{CT}(\nu)^\text{int} \in [0, t^\text{CT}], \nu \in \tilde{\Gamma}^\text{CT}_{ac}, ac^\text{CT}(\nu)^\text{int} \in [0, t^\text{CT}], \nu \in \tilde{\Gamma}^\text{CT}_{ac}:\) the number of edges \(e^\text{CT} \in E^\text{CT}\) (resp., edges \(e^\text{CT} \in E^\text{CT}\) with adjacency-configuration \(\nu\);
- \(ac^\text{FF}(\nu)^\text{int} \in [0, m^\text{FF}], \nu \in \tilde{\Gamma}^\text{FF}_{ac}, ac^\text{FF}(\nu)^\text{int} \in [0, m^\text{FF}], \nu \in \tilde{\Gamma}^\text{FF}_{ac}, ac^\text{FF}(\nu)^\text{int} \in [0, m^\text{FF}], \nu \in \tilde{\Gamma}^\text{FF}_{ac}:\) the number of edges \(e^\text{FF} \in E^\text{FF}\) with adjacency-configuration \(\nu\);
- \(ac^\text{CT}(\nu)^\text{int} \in [0, m^\text{CT}], \nu \in \tilde{\Gamma}^\text{CT}_{ac}, ac^\text{CT}(\nu)^\text{int} \in [0, m^\text{CT}], \nu \in \tilde{\Gamma}^\text{CT}_{ac}, ac^\text{FF}(\nu)^\text{int} \in [0, m^\text{FF}], \nu \in \tilde{\Gamma}^\text{FF}_{ac}, ac^\text{FF}(\nu)^\text{int} \in [0, m^\text{FF}], \nu \in \tilde{\Gamma}^\text{FF}_{ac}:\) the number of edges \(e^\text{CT} \in E^\text{CT}\) (resp., edges \(e^\text{CT} \in E^\text{CT}\) with adjacency-configuration \(\nu\);
- \(\delta^\text{FC}_{ac}(i, [\nu]^\text{int}) \in [0, 1], i \in [k^\text{FC} + 1, m^\text{FC}], \nu \in \tilde{\Gamma}^\text{FC}_{ac}, \delta^\text{FC}_{ac}(i, [\nu]^\text{int}) \in [0, 1], i \in [k^\text{FC} + 1, m^\text{FC}], \nu \in \tilde{\Gamma}^\text{FC}_{ac}:\) \(\delta^\text{FC}_{ac}(i, [\nu]^\text{int}) \in 1 \iff\) edge \(e^\text{X}_i\) has adjacency-configuration \(\nu\);
- $\delta_{ac}^{CT}(k, \nu) \in [0, 1], k \in [1, k_C] = I_{\geq 2} \cup I_{\geq 1}, \nu \in \Gamma_{ac}^{CT}$: $\delta_{ac}^{CT}(k, \nu) = 1$ (resp., $\delta_{ac}^{TC}(k, \nu) = 1$) $\Leftrightarrow$ edge $e_{tail}^{CT}(k, j)$ (resp., $e_{head}^{TC}(k, j)$) for some $j \in [1, t_T]$ has adjacency-configuration $\nu$;

- $\delta_{ac}^{CF}(c, \nu) \in [0, 1], c \in [1, t_C], \nu \in \Gamma_{ac}^{CF}$: $\delta_{ac}^{CF}(c, \nu) = 1$ $\Leftrightarrow$ edge $e_{c,i}^{CF}$ for some $i \in [1, t_F]$ has adjacency-configuration $\nu$;

- $\delta_{ac}^{TF}(i, \nu) \in [0, 1], i \in [1, t_T], \nu \in \Gamma_{ac}^{TF}$: $\delta_{ac}^{TF}(i, \nu) = 1$ $\Leftrightarrow$ edge $e_{i,j}^{TF}$ for some $j \in [1, t_F]$ has adjacency-configuration $\nu$;

- $\alpha_{ac}^{CT}(k, \alpha_{ac}^{TC}(k)) \in [0, 1], k \in [1, k_C]: \alpha(v)$ of the edge $(v_{tail}^{CT}(k), v) \in E_{CT}$ (resp., $(v, v_{head}^{CT}(k)) \in E_{TC}$) if any;

- $\alpha_{ac}^{CF}(c) \in [0, 1], c \in [1, t_C]: \alpha(v)$ of the edge $(v_{c}^{CF}, v) \in E_{CF}$ if any;

- $\alpha_{ac}^{TF}(i) \in [0, 1], i \in [1, t_T]: \alpha(v)$ of the edge $(v_{i}^{TF}, v) \in E_{TF}$ if any;

- $\Delta_{ac}^{CT+}(i), \Delta_{ac}^{CT-}(i) \in [0, 1], i \in [1, k_C]: \Delta_{ac}^{CT+}(i) = 0$ (resp., $\Delta_{ac}^{CT-}(i) = 0$) $\Leftrightarrow$ edge $e_{i,j}^{X} = (u, v) \in E_{X}$ is used in $G$ (resp., $e_{i,j}^{X} \not\in E(G)$);

- $\Delta_{ac}^{CT+}(k), \Delta_{ac}^{CT-}(k) \in [0, 1], k \in [1, k_C] = I_{\geq 2} \cup I_{\geq 1}; \Delta_{ac}^{CT+}(k) = \Delta_{ac}^{CT-}(k) = 0$ (resp., $\Delta_{ac}^{CT+}(k) = \alpha(u)$ and $\Delta_{ac}^{CT-}(k) = \alpha(v)$) $\Leftrightarrow$ edge $e_{tail}^{CT}(k, j) = (u, v) \in E_{CT}$ for some $j \in [1, t_T]$ is used in $G$ (resp., otherwise);

- $\Delta_{ac}^{TC+}(k), \Delta_{ac}^{TC-}(k) \in [0, 1], k \in [1, k_C] = I_{\geq 2} \cup I_{\geq 1};$ Analogous with $\Delta_{ac}^{CT+}(k)$ and $\Delta_{ac}^{CT-}(k)$;

- $\Delta_{ac}^{CF+}(c) \in [0, 1], \Delta_{ac}^{CF-}(c) \in [0, 1], c \in [1, t_C]: \Delta_{ac}^{CF+}(c) = \Delta_{ac}^{CF-}(c) = 0$ (resp., $\Delta_{ac}^{CF+}(c) = \alpha(u)$ and $\Delta_{ac}^{CF-}(c) = \alpha(v)$) $\Leftrightarrow$ edge $e_{c,i}^{CF} = (u, v) \in E_{CF}$ for some $i \in [1, t_F]$ is used in $G$ (resp., otherwise);

- $\Delta_{ac}^{TF+}(i) \in [0, 1], \Delta_{ac}^{TF-}(i) \in [0, 1], i \in [1, t_T];$ Analogous with $\Delta_{ac}^{CF+}(c)$ and $\Delta_{ac}^{CF-}(c)$;

constraints:

\begin{align*}
\text{acc}([\nu]_{\text{int}}) &= 0, \quad \nu \in \Gamma_{ac}^{\text{int}} \setminus \Gamma_{ac}^{\text{CF}}, \\
\text{act}([\nu]_{\text{int}}) &= 0, \quad \nu \in \Gamma_{ac}^{\text{int}} \setminus \Gamma_{ac}^{\text{TC}}, \\
\text{acf}([\nu]_{\text{int}}) &= 0, \quad \nu \in \Gamma_{ac}^{\text{int}} \setminus \Gamma_{ac}^{\text{TF}}, \\
\text{acCT}([\nu]_{\text{int}}) &= 0, \quad \nu \in \Gamma_{ac}^{\text{int}} \setminus \Gamma_{ac}^{\text{CT}}, \\
\text{acTC}([\nu]_{\text{int}}) &= 0, \quad \nu \in \Gamma_{ac}^{\text{int}} \setminus \Gamma_{ac}^{\text{AC}}, \\
\text{acCF}([\nu]_{\text{int}}) &= 0, \quad \nu \in \Gamma_{ac}^{\text{int}} \setminus \Gamma_{ac}^{\text{CF}}, \\
\text{actF}([\nu]_{\text{int}}) &= 0, \quad \nu \in \Gamma_{ac}^{\text{int}} \setminus \Gamma_{ac}^{\text{TF}},
\end{align*}

(62)
\[
\sum_{(a,b,m)\in \Gamma_{ac}^{\text{int}}} \alpha_{ac}([\nu]^{\text{int}}) = \sum_{i\in [k_{ac}+1,m_c]} \delta_{ac}^C(i,m), \quad m \in [1,3],
\]

\[
\sum_{(a,b,m)\in \Gamma_{ac}^{\text{int}}} \alpha_{T}([\nu]^{\text{int}}) = \sum_{i\in [2,t_T]} \delta_{ac}^T(i,m), \quad m \in [1,3],
\]

\[
\sum_{(a,b,m)\in \Gamma_{ac}^{\text{int}}} \alpha_{F}([\nu]^{\text{int}}) = \sum_{i\in [2,t_F]} \delta_{ac}^F(i,m), \quad m \in [1,3],
\]

\[
\sum_{(a,b,m)\in \Gamma_{ac}^{\text{int}}} \alpha_{CT}([\nu]^{\text{int}}) = \sum_{k\in [1,k_C]} \delta_{ac}^+(k,m), \quad m \in [1,3],
\]

\[
\sum_{(a,b,m)\in \Gamma_{ac}^{\text{int}}} \alpha_{TC}([\nu]^{\text{int}}) = \sum_{k\in [1,k_C]} \delta_{ac}^-(k,m), \quad m \in [1,3],
\]

\[
\sum_{(a,b,m)\in \Gamma_{ac}^{\text{int}}} \alpha_{CF}([\nu]^{\text{int}}) = \sum_{c\in [1,c]} \delta_{ac}^{-\text{int}}(c,m), \quad m \in [1,3],
\]

\[
\sum_{(a,b,m)\in \Gamma_{ac}^{\text{int}}} \alpha_{TF}([\nu]^{\text{int}}) = \sum_{c\in [\tilde{c}+1,c]} \delta_{ac}^{\text{int}}(c,m), \quad m \in [1,3],
\]

\[
\sum_{\nu=(a,b,m)\in \tilde{\Gamma}_{ac}^C} m \cdot \delta_{ac}^C(i,[\nu]^{\text{int}}) = \beta_{ac}^C(i), \quad \nu \in \tilde{\Gamma}_{ac}^C.
\]

\[
\Delta_{ac}^{C^+}(i) + \sum_{\nu=(a,b,m)\in \Gamma_{ac}^{\text{int}}} [a]^{\text{int}} \delta_{ac}^C(i,[\nu]^{\text{int}}) = \alpha_{ac}^C(\text{tail}(i)),
\]

\[
\Delta_{ac}^{C^-}(i) + \sum_{\nu=(a,b,m)\in \Gamma_{ac}^{\text{int}}} [b]^{\text{int}} \delta_{ac}^C(i,[\nu]^{\text{int}}) = \alpha_{ac}^C(\text{head}(i)), \quad i \in [k_{ac}+1,m_c],
\]

\[
\Delta_{ac}^{C^+}(i) + \Delta_{ac}^{C^-}(i) \leq 2|\Lambda^{\text{int}}|(1 - e_{ac}^C(i)), \quad \nu \in \tilde{\Gamma}_{ac}^C, \quad i \in [k_{ac}+1,m_c],
\]

\[
\sum_{i\in [k_{ac}+1,m_c]} \delta_{ac}^C(i,[\nu]^{\text{int}}) = \alpha_{ac}^C([\nu]^{\text{int}}), \quad \nu \in \tilde{\Gamma}_{ac}^C.
\]

\[
\sum_{\nu=(a,b,m)\in \tilde{\Gamma}_{ac}^{T}} m \cdot \delta_{ac}^T(i,[\nu]^{\text{int}}) = \beta_{ac}^T(i), \quad \nu \in \tilde{\Gamma}_{ac}^{T}.
\]

\[
\Delta_{ac}^{T^+}(i) + \sum_{\nu=(a,b,m)\in \Gamma_{ac}^{\text{int}}} [a]^{\text{int}} \delta_{ac}^T(i,[\nu]^{\text{int}}) = \alpha_{ac}^T(i-1),
\]

\[
\Delta_{ac}^{T^-}(i) + \sum_{\nu=(a,b,m)\in \Gamma_{ac}^{\text{int}}} [b]^{\text{int}} \delta_{ac}^T(i,[\nu]^{\text{int}}) = \alpha_{ac}^T(i), \quad i \in [2,t_T],
\]

\[
\Delta_{ac}^{T^+}(i) + \Delta_{ac}^{T^-}(i) \leq 2|\Lambda^{\text{int}}|(1 - e_{ac}^T(i)), \quad \nu \in \tilde{\Gamma}_{ac}^{T}, \quad i \in [2,t_T],
\]

\[
\sum_{i\in [2,t_T]} \delta_{ac}^T(i,[\nu]^{\text{int}}) = \alpha_{T}([\nu]^{\text{int}}), \quad \nu \in \tilde{\Gamma}_{ac}^{T}.
\]
\[
\sum_{\nu=(a,b,m)\in\Gamma^F_{ac}} m \cdot \delta^F_{ac}(i, [\nu]^{\text{int}}) = \beta^F(i),
\]
\[
\Delta^F_{ac}(i) + \sum_{\nu=(a,b,m)\in\Gamma^F_{ac}} [a]^{\text{int}} \delta^F_{ac}(i, [\nu]^{\text{int}}) = \alpha^F(i - 1),
\]
\[
\Delta^F_{ac}(i) + \sum_{\nu=(a,b,m)\in\Gamma^F_{ac}} [b]^{\text{int}} \delta^F_{ac}(i, [\nu]^{\text{int}}) = \alpha^F(i),
\]
\[
\Delta^F_{ac}(i) + \Delta^F_{ac}(i) \leq 2|\Lambda^\text{ext}|(1 - e^F(i)), \quad i \in [2, t_F],
\]
\[
\sum_{i\in[2,t_F]} \delta^F_{ac}(i, [\nu]^{\text{int}}) = ac_F([\nu]^{\text{int}}), \quad \nu \in \bar{\Gamma}^F_{ac},
\]
\[
\alpha^T(i) + |\Lambda^{\text{int}}|(1 - \chi^T(i, k) + e^T(i)) \geq \alpha^{CT}(k), \quad i \in [1, t_T],
\]
\[
\alpha^{CT}(k) \geq \alpha^T(i) - |\Lambda^{\text{int}}|(1 - \chi^T(i, k) + e^T(i)), \quad \sum_{\nu=(a,b,m)\in\Gamma^{CT}_{ac}} m \cdot \delta^{CT}_{ac}(k, [\nu]^{\text{int}}) = \beta^+(k),
\]
\[
\Delta^{CT}_{ac}(k) + \sum_{\nu=(a,b,m)\in\Gamma^{CT}_{ac}} [a]^{\text{int}} \delta^{CT}_{ac}(k, [\nu]^{\text{int}}) = \alpha^{CT}(k),
\]
\[
\Delta^{CT}_{ac}(k) + \sum_{\nu=(a,b,m)\in\Gamma^{CT}_{ac}} [b]^{\text{int}} \delta^{CT}_{ac}(k, [\nu]^{\text{int}}) = \alpha^{CT}(k),
\]
\[
\Delta^{CT}_{ac}(k) + \Delta^{CT}_{ac}(k) \leq 2|\Lambda^{\text{int}}|(1 - \delta^T(k)), \quad k \in [1, k_C],
\]
\[
\sum_{k\in[1,k_C]} \delta^{CT}_{ac}(k, [\nu]^{\text{int}}) = ac^{CT}([\nu]^{\text{int}}), \quad \nu \in \bar{\Gamma}^{CT}_{ac},
\]
\[
\alpha^T(i) + |\Lambda^{\text{int}}|(1 - \chi^T(i, k) + e^T(i + 1)) \geq \alpha^{TC}(k), \quad i \in [1, t_T],
\]
\[
\alpha^{TC}(k) \geq \alpha^T(i) - |\Lambda^{\text{int}}|(1 - \chi^T(i, k) + e^T(i + 1)), \quad \sum_{\nu=(a,b,m)\in\Gamma^{TC}_{ac}} m \cdot \delta^{TC}_{ac}(k, [\nu]^{\text{int}}) = \beta^-(k),
\]
\[
\Delta^{TC}_{ac}(k) + \sum_{\nu=(a,b,m)\in\Gamma^{TC}_{ac}} [a]^{\text{int}} \delta^{TC}_{ac}(k, [\nu]^{\text{int}}) = \alpha^{TC}(k),
\]
\[
\Delta^{TC}_{ac}(k) + \sum_{\nu=(a,b,m)\in\Gamma^{TC}_{ac}} [b]^{\text{int}} \delta^{TC}_{ac}(k, [\nu]^{\text{int}}) = \alpha^{C}(\text{head}(k)),
\]
\[
\Delta^{TC}_{ac}(k) + \Delta^{TC}_{ac}(k) \leq 2|\Lambda^{\text{int}}|(1 - \delta^T(k)), \quad k \in [1, k_C],
\]
\[
\sum_{k\in[1,k_C]} \delta^{TC}_{ac}(k, [\nu]^{\text{int}}) = ac^{TC}([\nu]^{\text{int}}), \quad \nu \in \bar{\Gamma}^{TC}_{ac},
\]
\[ \alpha^F (i) + |\Lambda^{\text{int}}| (1 - \chi^F (i, c) + e^F (i)) \geq \alpha^C (i), \]
\[ \alpha^C (c) \geq \alpha^F (i) - |\Lambda^{\text{int}}| (1 - \chi^F (i, c) + e^F (i)), \quad i \in [1, t_F], \]
\[ \sum_{\nu = (a, b, m) \in \Gamma^{\text{CF}}_{ac}} m \cdot \delta^\nu_{ac} (c, [\nu]^{\text{int}}) = \beta^{\text{in}} (c), \]
\[ \Delta^+_{ac} (c) + \sum_{\nu = (a, b, m) \in \Gamma^{\text{CF}}_{ac}} [a]^{\text{int}} \delta^\nu_{ac} (c, [\nu]^{\text{int}}) = \alpha^C (\text{head}(c)), \]
\[ \Delta_{ac} (c) + \sum_{\nu = (a, b, m) \in \Gamma^{\text{CF}}_{ac}} [b]^{\text{int}} \delta^\nu_{ac} (c, [\nu]^{\text{int}}) = \alpha^C (c), \]
\[ \Delta^+_{ac} (c) + \Delta^+_{ac} (c) \leq 2 \max \{|\Lambda^{\text{int}}|, |\Lambda^{\text{int}}|\} (1 - \delta^C (c)), \quad c \in [1, t_C], \]
\[ \sum_{c \in [1, t_C]} \delta^\nu_{ac} (c, [\nu]^{\text{int}}) = \alpha^{\text{CF}} ([\nu]^{\text{int}}), \quad \nu \in \Gamma^{\text{CF}}_{ac}, \quad (69) \]

\[ \alpha^F (j) + |\Lambda^{\text{int}}| (1 - \chi^F (j, i + t_C) + e^F (j)) \geq \alpha^T (j), \]
\[ \alpha^T (j) \geq \alpha^F (j) - |\Lambda^{\text{int}}| (1 - \chi^F (j, i + t_C) + e^F (j)), \quad j \in [1, t_F], \]
\[ \sum_{\nu = (a, b, m) \in \Gamma^{\text{TF}}_{ac}} m \cdot \delta^\nu_{ac} (i, [\nu]^{\text{int}}) = \beta^{\text{in}} (i + t_C), \]
\[ \Delta^+_{ac} (i) + \sum_{\nu = (a, b, m) \in \Gamma^{\text{TF}}_{ac}} [a]^{\text{int}} \delta^\nu_{ac} (i, [\nu]^{\text{int}}) = \alpha^T (i), \]
\[ \Delta_{ac} (i) + \sum_{\nu = (a, b, m) \in \Gamma^{\text{TF}}_{ac}} [b]^{\text{int}} \delta^\nu_{ac} (i, [\nu]^{\text{int}}) = \alpha^T (i), \]
\[ \Delta^+_{ac} (i) + \Delta^+_{ac} (i) \leq 2 \max \{|\Lambda^{\text{int}}|, |\Lambda^{\text{int}}|\} (1 - \delta^T (i + t_C)), \quad i \in [1, t_T], \]
\[ \sum_{i \in [1, t_T]} \delta^\nu_{ac} (i, [\nu]^{\text{int}}) = \alpha^{\text{TF}} ([\nu]^{\text{int}}), \quad \nu \in \Gamma^{\text{TF}}_{ac}, \quad (70) \]

\[ \sum_{\chi \in \{C, T, CF, CT, TC, CF\}} \alpha^{\chi} ([\nu]^{\text{int}}) + ac^\chi ([\overline{\nu}]^{\text{int}}) = \alpha^{\text{CF}} ([\nu]^{\text{int}}), \quad \nu \in \Gamma^{\text{int}}_{ac, <}, \]
\[ \sum_{\chi \in \{C, T, CF, CT, TC, CF\}} \alpha^{\chi} ([\nu]^{\text{int}}) = \alpha^{\text{CF}} ([\nu]^{\text{int}}), \quad \nu \in \Gamma^{\text{int}}_{ac, =}, \quad (71) \]

### 3.9 Descriptor for the Number of Chemical Symbols

We include constraints for computing the frequency of each chemical symbol in \( \Lambda_{dg} \). Let \( cs(v) \) denote the chemical symbol of a vertex \( v \) in a chemical graph \( G \) to be inferred; i.e., \( cs(v) = \mu = ad \in \Lambda_{dg} \) such that \( \alpha(v) = a \) and \( \text{deg}_G(v) = d \).

**constants:**

- A set \( \Lambda_{dg}^{\text{int}} \) of chemical symbols;
- Prepare a coding of each of the two sets \( \Lambda_{dg}^{\text{int}} \) and let \( [\mu]^{\text{int}} \) denote the coded integer of an element \( \mu \in \Lambda_{dg}^{\text{int}} \);
- Choose subsets \( \Lambda^C_{dg}, \Lambda^T_{dg}, \Lambda^F_{dg} \subseteq \Lambda^\text{int}_{dg} \): To compute the frequency of chemical symbols exactly, set \( \Lambda^C_{dg} := \Lambda^T_{dg} := \Lambda^F_{dg} := \Lambda^\text{int}_{dg} \).

variables:
- \( \text{ns}^{\text{int}}([\mu]^{\text{int}}) \in [0, n^{\text{int}}_{UB}], \mu \in \Lambda^\text{int}_{dg} \): the number of interior-vertices \( v \) with \( \text{cs}(v) = \mu \);
- \( \delta_{\text{ns}}^X(i, [\mu]^{\text{int}}) \in [0, 1], i \in [1, t_X], \mu \in \Lambda^\text{int}_{dg}, X \in \{ C, T, F \} \);

constraints:
\[
\sum_{\mu \in \Lambda^X_{dg}(e)} \delta_{\text{ns}}^X(i, [\mu]^{\text{int}}) = 1, \quad \sum_{\mu = ad \in \Lambda^X_{dg}} [a]^{\text{int}} \cdot \delta_{\text{ns}}^X(i, [\mu]^{\text{int}}) = \alpha^X(i),
\]
\[
\sum_{\mu = ad \in \Lambda^X_{dg}} d \cdot \delta_{\text{ns}}^X(i, [\mu]^{\text{int}}) = \text{deg}^X(i),
\]
\[
i \in [1, t_X], X \in \{ C, T, F \},
\]
\[
\sum_{i \in [1, t_c]} \delta_{\text{ns}}^C(i, [\mu]^{\text{int}}) + \sum_{i \in [1, t_t]} \delta_{\text{ns}}^T(i, [\mu]^{\text{int}}) + \sum_{i \in [1, t_f]} \delta_{\text{ns}}^F(i, [\mu]^{\text{int}}) = \text{ns}^{\text{int}}([\mu]^{\text{int}}), \quad \mu \in \Lambda^\text{int}_{dg}.
\]

3.10 Descriptor for the Number of Edge-configurations

We include constraints to compute the frequency of each edge-configuration in an inferred chemical graph \( G \).

constants:
- A set \( \Gamma^\text{int} \) of edge-configurations \( \gamma = (\mu, \xi, m) \) with \( \mu \leq \xi \);
- Let \( \Gamma^\text{int}_< = \{ (\mu, \xi, m) \in \Gamma^\text{int} \mid \mu < \xi \} \), \( \Gamma^\text{int}_\leq = \{ (\mu, \xi, m) \in \Gamma^\text{int} \mid \mu = \xi \} \) and \( \Gamma^\text{int}_\geq = \{ (\xi, \mu, m) \mid (\mu, \xi, m) \in \Gamma^\text{int}_< \} \);
- Prepare a coding of the set \( \Gamma^\text{int}_\cup \Gamma^\text{int}_\geq \) and let \([\gamma]^{\text{int}}\) denote the coded integer of an element \( \gamma \) in \( \Gamma^\text{int}_\cup \Gamma^\text{int}_\geq \);
- Choose subsets \( \Gamma^C_{ec}, \Gamma^T_{ec}, \Gamma^F_{ec}, \Gamma^T_{ec}, \Gamma^C_{ec}, \Gamma^T_{ec}, \Gamma^F_{ec} \subseteq \Gamma^\text{int}_\cup \Gamma^\text{int}_\geq \): To compute the frequency of edge-configurations exactly, set \( \Gamma^\text{C}_{ec} := \Gamma^T_{ec} := \Gamma^C_{ec} := \Gamma^T_{ec} := \Gamma^F_{ec} := \Gamma^T_{ec} := \Gamma^F_{ec} := \Gamma^\text{int}_\cup \Gamma^\text{int}_\geq \);
- \( \text{ec}_{\text{LB}}(\gamma), \text{ec}_{\text{UB}}(\gamma) \in [0, 2n^{\text{int}}_{UB}], \gamma = (\mu, \xi, m) \in \Gamma^\text{int}_\cup \Gamma^\text{int}_\geq \): lower and upper bounds on the number of interior-edges \( e = uv \) with \( \text{cs}(u) = \mu, \text{cs}(v) = \xi \) and \( \beta(e) = m \);

variables:
- \( \text{ec}^{\text{int}}([\gamma]^{\text{int}}) \in [\text{ec}_{\text{LB}}(\gamma), \text{ec}_{\text{UB}}(\gamma)], \gamma \in \Gamma^\text{int}_\cup \Gamma^\text{int}_\geq \): the number of interior-edges with edge-configuration \( \gamma \);
- \( \text{ec}^{\text{C}}([\gamma]^{\text{int}}) \in [0, m_C], \gamma \in \Gamma^\text{C}_{ec} \), \( \text{ec}^{\text{T}}([\gamma]^{\text{int}}) \in [0, t_T], \gamma \in \Gamma^\text{T}_{ec} \), \( \text{ec}^{\text{F}}([\gamma]^{\text{int}}) \in [0, t_F], \gamma \in \Gamma^\text{F}_{ec} \): the number of edges \( e^C \in E_C \) (resp., edges \( e^T \in E_T \) and edges \( e^F \in E_F \)) with edge-configuration \( \gamma \);
- \( \text{ec}^{\text{CT}}([\gamma]^{\text{int}}) \in [0, \min\{k_C, t_T\}], \gamma \in \Gamma^\text{CT}_{ec}, \text{ec}^{\text{TC}}([\gamma]^{\text{int}}) \in [0, \min\{k_T, t_T\}], \gamma \in \Gamma^\text{TC}_{ec} \), \( \text{ec}^{\text{CF}}([\gamma]^{\text{int}}) \in [0, t_C], \gamma \in \Gamma^\text{CF}_{ec}, \text{ec}^{\text{TF}}([\gamma]^{\text{int}}) \in [0, t_T], \gamma \in \Gamma^\text{TF}_{ec} \): the number of edges \( e^{CT} \in E_{CT} \) (resp., edges \( e^{TC} \in E_{TC} \)) and edges \( e^{CF} \in E_{CF} \) and \( e^{TF} \in E_{TF} \)) with edge-configuration \( \gamma \).
- $\delta^C_{ec}(i, [\gamma]^{int}) \in [0, 1], i \in [\widetilde{k}_C + 1, m_C] = I_{(\geq 1)} \cup I_{(0/1)} \cup I_{(= 1)}, \gamma \in \Gamma_{ec}^C; \delta^T_{ec}(i, [\gamma]^{int}) \in [0, 1], i \in [2, t_T], \gamma \in \Gamma_{ec}^T, \delta_{ec}^F(i, [\gamma]^{int}) \in [0, 1], i \in [2, t_F], \gamma \in \Gamma_{ec}^F$: $\delta_{ec}^{\gamma}(i, [\gamma]^{1}) = 1$ \iff edge $e^X_{i}$ has edge-configuration $\gamma$;

- $\delta_{ec,C}^{CT}(k, [\gamma]^{int}), \delta_{ec,C}^{TC}(k, [\gamma]^{int}) \in [0, 1], k \in [1, k_C] = I_{(\geq 2)} \cup I_{(\geq 1)}, \gamma \in \Gamma_{ec}^{CT}; \delta_{ec,C}^{CT}(k, [\gamma]^{int}) = 1$ (resp., $\delta_{ec,C}^{TC}(k, [\gamma]^{int}) = 1$) \iff edge $e^{CT\_tail(k)}(j)$ (resp., $e^{TC\_head(k)}(j)$) for some $j \in [1, t_T]$ has edge-configuration $\gamma$;

- $\delta_{ec,C}^{CF}(c, [\gamma]^{int}) \in [0, 1], c \in [1, \widetilde{r}_C], \gamma \in \Gamma_{ec}^{CF}; \delta_{ec,C}^{CF}(c, [\gamma]^{int}) = 1 \iff edge e^{CF\_ci}$ for some $i \in [1, t_F]$ has edge-configuration $\gamma$;

- $\delta_{ec,T}^{TF}(i, [\gamma]^{int}) \in [0, 1], i \in [1, t_T], \gamma \in \Gamma_{ec}^{TF}; \delta_{ec,T}^{TF}(i, [\gamma]^{int}) = 1 \iff edge e^{TF\_i,j}$ for some $j \in [1, t_F]$ has edge-configuration $\gamma$;

- $\deg_{CT}^{CT}(k), \deg_{CT}^{TC}(k) \in [0, 4], k \in [1, k_C]: \deg_G(v) of an end-vertex $v \in V_T$ of the edge $(v^{C\_tail(k)}, v) \in E_{CT}$ (resp., $(v, v^{C\_head(k)}) \in E_{TC}$) if any;

- $\deg_{TF}^{CF}(c) \in [0, 4], c \in [1, \widetilde{r}_C]: \deg_G(v) of an end-vertex $v \in V_F$ of the edge $(v^{C\_v}, v) \in E_{CF}$ if any;

- $\deg_{TF}^{TF}(i) \in [0, 4], i \in [1, t_T]: \deg_G(v) of an end-vertex $v \in V_F$ of the edge $(v^{T\_i}, v) \in E_{TF}$ if any;

- $\Delta_{ec}^{C+}(i), \Delta_{ec}^{C-}(i) \in [0, 4], i \in [\widetilde{k}_C + 1, m_C], \Delta_{ec}^{CT+}(i), \Delta_{ec}^{CT-}(i) \in [0, 4], i \in [2, t_T], \Delta_{ec}^{TF+}(i), \Delta_{ec}^{TF-}(i) \in [0, 4], i \in [2, t_F]: \Delta_{ec}^{X+}(i) = \Delta_{ec}^{X-}(i) = 0$ (resp., $\Delta_{ec}^{X+}(i) = \deg_G(u) and \Delta_{ec}^{X-}(i) = \deg_G(v)$) \iff edge $e^{X\_i,(u,v)} \in E_X$ is used in $G$ (resp., $e^{X\_i} \notin E(G)$);

- $\Delta_{ec}^{CT+}(k), \Delta_{ec}^{CT-}(k) \in [0, 4], k \in [1, k_C] = I_{(\geq 2)} \cup I_{(\geq 1)}$: $\Delta_{ec}^{CT+}(k) = \Delta_{ec}^{CT-}(k) = 0$ (resp., $\Delta_{ec}^{CT+}(k) = \deg_G(u) and \Delta_{ec}^{CT-}(k) = \deg_G(v)$) \iff edge $e^{CT\_tail(k),j} = (u,v) \in E_{CT}$ for some $j \in [1, t_T]$ is used in $G$ (resp., otherwise);

- $\Delta_{ec}^{TC+}(k), \Delta_{ec}^{TC-}(k) \in [0, 4], k \in [1, k_C] = I_{(\geq 2)} \cup I_{(\geq 1)}$: Analogous with $\Delta_{ec}^{CT+}(k)$ and $\Delta_{ec}^{CT-}(k)$;

- $\Delta_{ec}^{CF+}(c), \Delta_{ec}^{CF-}(c) \in [0, 4], c \in [1, \widetilde{r}_C]: \Delta_{ec}^{CF+}(c) = \Delta_{ec}^{CF-}(c) = 0$ (resp., $\Delta_{ec}^{CF+}(c) = \deg_G(u) and \Delta_{ec}^{CF-}(c) = \deg_G(v)$) \iff edge $e^{CF\_cj} = (u,v) \in E_{CF}$ for some $j \in [1, t_F]$ is used in $G$ (resp., otherwise);

- $\Delta_{ec}^{TF+}(i), \Delta_{ec}^{TF-}(i) \in [0, 4], i \in [1, t_T]:$ Analogous with $\Delta_{ec}^{CF+}(c)$ and $\Delta_{ec}^{CF-}(c)$;

\textbf{constraints:}

\begin{align*}
\text{ecc}(\gamma) &= 0, \quad \gamma \in \Gamma^{int} \setminus \Gamma_{ec}^{C}, \\
\text{ec}_{T}(\gamma) &= 0, \quad \gamma \in \Gamma^{int} \setminus \Gamma_{ec}^{T}, \\
\text{ec}_{F}(\gamma) &= 0, \quad \gamma \in \Gamma^{int} \setminus \Gamma_{ec}^{F}, \\
\text{ec}_{CT}(\gamma) &= 0, \quad \gamma \in \Gamma^{int} \setminus \Gamma_{ec}^{CT}, \\
\text{ec}_{TC}(\gamma) &= 0, \quad \gamma \in \Gamma^{int} \setminus \Gamma_{ec}^{TC}, \\
\text{ec}_{CF}(\gamma) &= 0, \quad \gamma \in \Gamma^{int} \setminus \Gamma_{ec}^{CF}, \\
\text{ec}_{TF}(\gamma) &= 0, \quad \gamma \in \Gamma^{int} \setminus \Gamma_{ec}^{TF}, \\
\end{align*}
\[
\sum_{(\mu,\mu',m)=\gamma \in \Gamma^\text{int}} \mathcal{E}_C^\Gamma([\gamma])^{\text{int}} = \sum_{i \in [k_C+1,m_C]} \gamma_{\beta}^C(i,m), \quad m \in [1,3],
\]
\[
\sum_{(\mu,\mu',m)=\gamma \in \Gamma_T^\text{int}} \mathcal{E}_T^\Gamma([\gamma])^{\text{int}} = \sum_{i \in [2,t_T]} \gamma_{\beta}^T(i,m), \quad m \in [1,3],
\]
\[
\sum_{(\mu,\mu',m)=\gamma \in \Gamma_F^\text{int}} \mathcal{E}_F^\Gamma([\gamma])^{\text{int}} = \sum_{i \in [2,t_F]} \gamma_{\beta}^F(i,m), \quad m \in [1,3],
\]
\[
\sum_{(\mu,\mu',m)=\gamma \in \Gamma_T^\text{int}} \mathcal{E}_C^\Gamma([\gamma])^{\text{int}} = \sum_{k \in [1,k_C]} \delta_{\beta}^+(k,m), \quad m \in [1,3],
\]
\[
\sum_{(\mu,\mu',m)=\gamma \in \Gamma_T^\text{int}} \mathcal{E}_C^\Gamma([\gamma])^{\text{int}} = \sum_{k \in [1,k_C]} \delta_{\beta}^-(k,m), \quad m \in [1,3],
\]
\[
\sum_{(\mu,\mu',m)=\gamma \in \Gamma_T^\text{int}} \mathcal{E}_C^\Gamma([\gamma])^{\text{int}} = \sum_{c \in [1,c^+_C]} \delta_{\beta}^{c}(c,m), \quad m \in [1,3],
\]
\[
\sum_{(\mu,\mu',m)=\gamma \in \Gamma_T^\text{int}} \mathcal{E}_C^\Gamma([\gamma])^{\text{int}} = \sum_{c \in [c^+_C+1,c^+_T]} \delta_{\beta}^{c}(c,m), \quad m \in [1,3],
\]
(75)

\[
\sum_{\gamma=(a,b,m) \in \tilde{\Gamma}_C^g} \mathcal{E}_C^\Gamma([\gamma])^{\text{int}} \cdot \delta_{\text{ec}}^C(i,\gamma) = \sum_{\nu \in \tilde{\Gamma}_C^g} \mathcal{E}_C^\Gamma([\gamma])^{\text{int}} \cdot \delta_{\text{ac}}^C(i,\gamma),
\]
\[
\begin{aligned}
\Delta_{\text{ec}}^C(i) &+ \sum_{\gamma=(a,b,m) \in \tilde{\Gamma}_C^g} d \cdot \delta_{\text{ec}}^C(i,\gamma) = \deg^C(\text{tail}(i)), \\
\Delta_{\text{ec}}^{C^+}(i) &+ \sum_{\gamma=(a,b,m) \in \tilde{\Gamma}_C^g} d \cdot \delta_{\text{ec}}^C(i,\gamma) = \deg^C(\text{head}(i)), \\
\Delta_{\text{ec}}^{C^+}(i) + \Delta_{\text{ec}}^{C^-}(i) &\leq 8(1 - e^C(i)), \\
\sum_{i \in [k_C+1,m_C]} \delta_{\text{ec}}^C(i,\gamma) &\leq \mathcal{E}_C^\Gamma([\gamma])^{\text{int}}, \quad \gamma \in \tilde{\Gamma}_C^g,
\end{aligned}
\]  
(76)

\[
\sum_{\gamma=(a,b,m) \in \tilde{\Gamma}_T^g} \mathcal{E}_C^\Gamma([\gamma])^{\text{int}} \cdot \delta_{\text{ec}}^T(i,\gamma) = \sum_{\nu \in \tilde{\Gamma}_T^g} \mathcal{E}_C^\Gamma([\gamma])^{\text{int}} \cdot \delta_{\text{ac}}^T(i,\gamma),
\]
\[
\begin{aligned}
\Delta_{\text{ec}}^T(i) &+ \sum_{\gamma=(a,b,m) \in \tilde{\Gamma}_T^g} d \cdot \delta_{\text{ec}}^T(i,\gamma) = \deg^T(i-1), \\
\Delta_{\text{ec}}^{T^+}(i) &+ \sum_{\gamma=(a,b,m) \in \tilde{\Gamma}_T^g} d \cdot \delta_{\text{ec}}^T(i,\gamma) = \deg^T(i), \\
\Delta_{\text{ec}}^{T^+}(i) + \Delta_{\text{ec}}^{T^-}(i) &\leq 8(1 - e^T(i)), \\
\sum_{i \in [2,t_T]} \delta_{\text{ec}}^T(i,\gamma) &\leq \mathcal{E}_T^\Gamma([\gamma])^{\text{int}}, \quad \gamma \in \tilde{\Gamma}_T^g,
\end{aligned}
\]  
(77)
\[
\sum_{\gamma=(ad, bd', m) \in \Gamma_{ec}^{F}} [(a, b, m)]^{\text{int}} \cdot \delta_{ec}^{F}(i, [\gamma])^{\text{int}} = \sum_{\nu \in \Gamma_{ac}^{F}} [\nu]^{\text{int}} \cdot \delta_{ac}^{F}(i, [\nu])^{\text{int}},
\]

\[
\Delta_{ec}^{F+}(i) + \sum_{\gamma=(ad, \xi, m) \in \Gamma_{ec}^{F}} d \cdot \delta_{ec}^{F}(i, [\gamma])^{\text{int}} = \text{deg}^{F}(i - 1),
\]

\[
\Delta_{ec}^{F-}(i) + \sum_{\gamma=(\mu, bd, m) \in \Gamma_{ec}^{F}} d \cdot \delta_{ec}^{F}(i, [\gamma])^{\text{int}} = \text{deg}^{F}(i),
\]

\[
\Delta_{ec}^{F+}(i) + \Delta_{ec}^{F-}(i) \leq 8(1 - e^{F}(i)), \quad i \in [2, t_{F}],
\]

\[
\sum_{i \in [2, t_{F}]} \delta_{ec}^{F}(i, [\gamma])^{\text{int}} = \text{ec}_{F}(\gamma)^{\text{int}}, \quad \gamma \in \Gamma_{ec}^{F},
\]

\[
\text{deg}^{T}(i) + 4(1 - \chi^{T}(i, k) + e^{T}(i)) \geq \text{deg}^{CT}(k),
\]

\[
\text{deg}^{CT}(k) \geq \text{deg}^{T}(i) - 4(1 - \chi^{T}(i, k) + e^{T}(i)), \quad i \in [1, t_{T}],
\]

\[
\sum_{\gamma=(ad, bd', m) \in \Gamma_{ec}^{CT}} [(a, b, m)]^{\text{int}} \cdot \delta_{ec, C}^{CT}(k, [\gamma])^{\text{int}} = \sum_{\nu \in \Gamma_{ac}^{CT}} [\nu]^{\text{int}} \cdot \delta_{ac}^{CT}(k, [\nu])^{\text{int}},
\]

\[
\Delta_{ec}^{CT+}(k) + \sum_{\gamma=(ad, \xi, m) \in \Gamma_{ec}^{CT}} d \cdot \delta_{ec, C}^{CT}(k, [\gamma])^{\text{int}} = \text{deg}^{CT}(k),
\]

\[
\Delta_{ec}^{CT-}(k) + \sum_{\gamma=(\mu, bd, m) \in \Gamma_{ec}^{CT}} d \cdot \delta_{ec, C}^{CT}(k, [\gamma])^{\text{int}} = \text{deg}^{CT}(k),
\]

\[
\Delta_{ec}^{CT+}(k) + \Delta_{ec}^{CT-}(k) \leq 8(1 - \delta_{T}^{T}(k)), \quad k \in [1, k_{C}],
\]

\[
\sum_{k \in [1, k_{C}]} \delta_{ec, C}(k, [\gamma])^{\text{int}} = \text{ec}_{CT}(\gamma)^{\text{int}}, \quad \gamma \in \Gamma_{ec}^{CT},
\]

\[
\text{deg}^{T}(i) + 4(1 - \chi^{T}(i, k) + e^{T}(i + 1)) \geq \text{deg}^{TC}(k),
\]

\[
\text{deg}^{TC}(k) \geq \text{deg}^{T}(i) - 4(1 - \chi^{T}(i, k) + e^{T}(i + 1)), \quad i \in [1, t_{T}],
\]

\[
\sum_{\gamma=(ad, bd', m) \in \Gamma_{ec}^{TC}} [(a, b, m)]^{\text{int}} \cdot \delta_{ec, C}^{TC}(k, [\gamma])^{\text{int}} = \sum_{\nu \in \Gamma_{ac}^{TC}} [\nu]^{\text{int}} \cdot \delta_{ac}^{TC}(k, [\nu])^{\text{int}},
\]

\[
\Delta_{ec}^{TC+}(k) + \sum_{\gamma=(ad, \xi, m) \in \Gamma_{ec}^{TC}} d \cdot \delta_{ec, C}^{TC}(k, [\gamma])^{\text{int}} = \text{deg}^{TC}(k),
\]

\[
\Delta_{ec}^{TC-}(k) + \sum_{\gamma=(\mu, bd, m) \in \Gamma_{ec}^{TC}} d \cdot \delta_{ec, C}^{TC}(k, [\gamma])^{\text{int}} = \text{deg}^{TC}(k),
\]

\[
\Delta_{ec}^{TC+}(k) + \Delta_{ec}^{TC-}(k) \leq 8(1 - \delta_{T}^{T}(k)), \quad k \in [1, k_{C}],
\]

\[
\sum_{k \in [1, k_{C}]} \delta_{ec, C}(k, [\gamma])^{\text{int}} = \text{ec}_{TC}(\gamma)^{\text{int}}, \quad \gamma \in \Gamma_{ec}^{TC},
\]
\[ \text{deg}^F(i) + 4(1 - x^F(i, c) + e^F(i)) \geq \text{deg}^F_G(c), \]
\[ \text{deg}^G_G(c) \geq \text{deg}^F(i) - 4(1 - x^F(i, c) + e^F(i)), \quad i \in [1, t_F], \]
\[ \sum_{\gamma=(a,d',m)} [(a, b, m)]^{\text{int}} \cdot \delta^C_{\text{ec}, C}(c, [\gamma])^{\text{int}} = \sum_{\nu \in \Gamma^C_{\text{ec}}} [\nu]^{\text{int}} \cdot \delta^C_{\text{ac}, C}(c, [\nu])^{\text{int}}, \]
\[ \Delta^C_{\text{ec}}+(c) + \sum_{\gamma=(a,d',m)} d \cdot \delta^C_{\text{ec}, C}(c, [\gamma])^{\text{int}} = \text{deg}^C(c), \]
\[ \Delta^C_{\text{ec}}-(c) + \sum_{\gamma=(a,d',m)} d \cdot \delta^C_{\text{ec}, C}(c, [\gamma])^{\text{int}} = \text{deg}^F_G(c), \]
\[ \Delta^C_{\text{ec}}+(c) + \Delta^C_{\text{ec}}-(c) \leq 8(1 - \delta^C_{\text{ec}}(c)), \quad c \in [1, t_C], \]
\[ \sum_{c \in [1, t_C]} \delta_{\text{ec}, C}(c, [\gamma])^{\text{int}} = \text{ec}_{\text{ec}}([\gamma])^{\text{int}}, \quad \gamma \in \Gamma^C_{\text{ec}}, \quad (81) \]
\[ \text{deg}^F(j) + 4(1 - x^F(j, i + t_C) + e^F(j)) \geq \text{deg}^F_G(i), \]
\[ \text{deg}^F_G(i) \geq \text{deg}^F(j) - 4(1 - x^F(j, i + t_C) + e^F(j)), \quad j \in [1, t_F], \]
\[ \sum_{\gamma=(a,d',m)} [(a, b, m)]^{\text{int}} \cdot \delta^F_{\text{ec}, T}(i, [\gamma])^{\text{int}} = \sum_{\nu \in \Gamma^F_{\text{ec}}} [\nu]^{\text{int}} \cdot \delta^F_{\text{ac}, T}(i, [\nu])^{\text{int}}, \]
\[ \Delta^F_{\text{ec}}+(i) + \sum_{\gamma=(a,d',m)} d \cdot \delta^F_{\text{ec}, T}(i, [\gamma])^{\text{int}} = \text{deg}^T_G(i), \]
\[ \Delta^F_{\text{ec}}-(i) + \sum_{\gamma=(a,d',m)} d \cdot \delta^F_{\text{ec}, T}(i, [\gamma])^{\text{int}} = \text{deg}^F_G(i), \]
\[ \Delta^F_{\text{ec}}+(i) + \Delta^F_{\text{ec}}-(i) \leq 8(1 - \delta^F_{\text{ec}}(i)), \quad i \in [1, t_T], \]
\[ \sum_{i \in [1, t_T]} \delta_{\text{ec}, T}(i, [\gamma])^{\text{int}} = \text{ec}_{\text{T}}([\gamma])^{\text{int}}, \quad \gamma \in \Gamma^F_{\text{ec}}, \quad (82) \]
\[ \sum_{X \in \{C, T, F, CT, TC, CF, TF\}} \text{ec}_X([\gamma])^{\text{int}} = \text{ec}_{\text{INT}}([\gamma])^{\text{int}}, \quad \gamma \in \Gamma^T_{\text{INT}}, \]
\[ \sum_{X \in \{C, T, F, CT, TC, CF, TF\}} \text{ec}_X([\gamma])^{\text{int}} = \text{ec}_{\text{INT}}([\gamma])^{\text{int}}, \quad \gamma \in \Gamma^F_{\text{INT}}, \quad (83) \]

### 3.11 Descriptor for the Number of of Fringe-configurations

We include constraints to compute the frequency of each fringe-configuration in an inferred chemical graph \( G \).

**variables:**

\( \text{fc}([\psi]) \in [0, t_C + t_T + t_F], \psi \in \mathcal{F}^{*} \): the frequency of a chemical rooted tree \( \psi \) in the set of \( \rho \)-fringe-trees in \( G \);

**constraints:**

...
3.12 Constraints for Normalization of Feature Vectors

By introducing a tolerance $\varepsilon > 0$ in the conversion between integers and reals, we include the following constraints for normalizing of a feature vector $f(G) = (x_1, x_2, \ldots, x_K)$:

\[
\frac{(1 - \varepsilon)(x_i - \min(dcp_i; D_\pi))}{\max(dcp_i; D_\pi) - \min(dcp_i; D_\pi)} \leq \hat{x}_i \leq \frac{(1 + \varepsilon)(x_i - \min(dcp_i; D_\pi))}{\max(dcp_i; D_\pi) - \min(dcp_i; D_\pi)}, \quad i \in [1, K].
\] (85)

An example of a tolerance is $\varepsilon = 0.01$.

References

[1] R. Ito, N. A. Azam, C. Wang, A. Shurbevski, H. Nagamochi, T. Akutsu, A novel method for the inverse QSAR/QSPR to monocyclic chemical compounds based on artificial neural networks and integer programming, BIOCOMP2020, Las Vegas, Nevada, USA, 27-30 July 2020.

[2] F. Zhang, J. Zhu, R. Chiewvanichakorn, A. Shurbevski, H. Nagamochi, T. Akutsu, A new integer linear programming formulation to the inverse QSAR/QSPR for acyclic chemical compounds using skeleton trees, The 33rd International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, September 22-25, 2020 Kitakyushu, Japan, Springer LNCS 12144, pp. 433–444.

[3] N. A. Azam, J. Zhu, Y. Sun, Y. Shi, A. Shurbevski, L. Zhao, H. Nagamochi, T. Akutsu, A novel method for inference of acyclic chemical compounds with bounded branch-height based on artificial neural networks and integer programming, arXiv:2009.09646

[4] T. Akutsu, H. Nagamochi, A novel method for inference of chemical compounds with prescribed topological substructures based on integer programming, arXiv: 2010.09203, 2020.