Supplementary Materials

A Method for Inferring Polymers Based on Linear Regression and Integer Programming

Ryota Ido[1], Shengjuan Cao[1], Jianshen Zhu[1], Naveed Ahmed Azam[2], Kazuya Haraguchi[1], Liang Zhao[3], Hiroshi Nagamochi[1], Tatsuya Akutsu[4]

[1]Department of Applied Mathematics and Physics, Kyoto University, Kyoto 606-8501, Japan;
[2]Department of Mathematics, Quaid-i-Azam University, Islamabad 45320, Pakistan;
[3]Graduate School of Advanced Integrated Studies in Human Survivability (Shishu-Kan), Kyoto University, Kyoto 606-8306, Japan;
[4]Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji 611-0011, Japan

1 Linear Regressions

This section reviews the method for linear regression used by Zhu et al. [1] in the framework of inferring chemical graphs.

For an integer $p \geq 1$ and a vector $x \in \mathbb{R}^p$, the j-th entry of x is denoted by $x(j), j \in [1, p]$.

Let D be a data set of chemical graphs $\mathbb{C}$ with an observed value $a(\mathbb{C}) \in \mathbb{R}$, where we denote by $a_i = a(\mathbb{C}_i)$ for an indexed graph $\mathbb{C}_i$.

Let f be a feature function that maps a chemical graph $\mathbb{C}$ to a vector $f(\mathbb{C}) \in \mathbb{R}^K$ where we denote by $x_i = f(\mathbb{C}_i)$ for an indexed graph $\mathbb{C}_i$. For a prediction function $\eta : \mathbb{R}^K \to \mathbb{R}$, define an error function

$$\mathrm{Err}(\eta; D) \triangleq \sum_{\mathbb{C}_i \in D} (a_i - \eta(f(\mathbb{C}_i)))^2 = \sum_{\mathbb{C}_i \in D} (a_i - \eta(x_i))^2,$$

and define the *coefficient of determination* $\mathrm{R}^2(\eta, D)$ to be

$$\mathrm{R}^2(\eta, D) \triangleq 1 - \frac{\mathrm{Err}(\eta; D)}{\sum_{\mathbb{C}_i \in D}(a_i - \widetilde{a})^2} \text{ for } \widetilde{a} = \frac{1}{|D|} \sum_{\mathbb{C} \in D} a(\mathbb{C}).$$

For a feature space $\mathbb{R}^K$, a hyperplane is defined to be a pair (w, b) of a vector $w \in \mathbb{R}^K$ and a real $b \in \mathbb{R}$. Given a hyperplane $(w, b) \in \mathbb{R}^{K+1}$, a prediction function $\eta_{w,b} : \mathbb{R}^K \to \mathbb{R}$ is defined by setting

$$\eta_{w,b}(x) \triangleq w \cdot x + b = \sum_{j \in [1,K]} w(j)x(j) + b.$$

We observe that such a prediction function can be represented as an ANN with an input layer with K nodes $u_j, j \in [1, K]$ and an output layer with a single node v such that the weight of edge arc (u_j, v) is set to be $w(j)$, the bias of node u is set to be b and the activation function at node u is set to be a linear function. However, a learning algorithm for an ANN may not find a set of weights $w(j), j \in [1, K]$ and b that minimizes the error function, since the algorithm simply iterates modification of the current weights and biases until it terminates at a local optima in the minimization.

We wish to find a hyperplane (w, b) that minimizes the error function $\mathrm{Err}(\eta_{w,b}; D)$. In many cases, a feature vector f contains descriptors that do not play an essential role in constructing a good prediction function. When we solve the minimization problem, the entries $w(j)$ for some descriptors $j \in [1, K]$ in the resulting hyperplane (w, b) become zero, which means that these descriptors were not necessarily important for finding a prediction function $\eta_{w,b}$. It is proposed that solving the minimization with an additional penalty term τ to the error function often results in a more number of entries $w(j) = 0$, reducing a set of descriptors necessary for defining a prediction function $\eta_{w,b}$. For an error function with such a penalty term, a Ridge function $\frac{1}{2|D|}\mathrm{Err}(\eta_{w,b}; D)+\lambda[\sum_{j\in[1,K]} w(j)^2+b^2]$ [2, 3] and a Lasso function $\frac{1}{2|D|}\mathrm{Err}(\eta_{w,b}; D) + \lambda[\sum_{j\in[1,K]} |w(j)| + |b|]$ [4] are known, where $\lambda \in \mathbb{R}$ is a given real number.

Given a prediction function $\eta_{w,b}$, we can simulate a process of computing the output $\eta_{w,b}(x)$ for an input $x \in \mathbb{R}^K$ as an MILP $\mathcal{M}(x, y; \mathcal{C}_1)$ in the framework. By solving such an MILP for a specified target value y^*, we can find a vector $x^* \in \mathbb{R}^K$ such that $\eta_{w,b}(x^*) = y^*$. Instead of specifying a single target value y^*, we use lower and upper bounds $\underline{y}^*, \overline{y}^* \in \mathbb{R}$ on the value $a(\mathbb{C})$ of a chemical graph $\mathbb{C}$ to be inferred. We can control the range between $\underline{y}^*$ and $\overline{y}^*$ for searching a chemical graph $\mathbb{C}$ by setting $\underline{y}^*$ and $\overline{y}^*$ to be close or different values. A desired MILP is formulated as follows.

$\mathcal{M}(x, y; \mathcal{C}_1)$**: An MILP formulation for the inverse problem to prediction function**

constants:
- A hyperplane (w, b) with $w \in \mathbb{R}^K$ and $b \in \mathbb{R}$;
- Real values $\underline{y}^*, \overline{y}^* \in \mathbb{R}$ such that $\underline{y}^* < \overline{y}^*$;
- A set $I_{\mathbb{Z}}$ of indices $j \in [1, K]$ such that the j-th descriptor $\mathrm{dcp}_j(\mathbb{C})$ is always an integer;
- A set I_+ of indices $j \in [1, K]$ such that the j-th descriptor $\mathrm{dcp}_j(\mathbb{C})$ is always non-negative;
- $\ell(j), u(j) \in \mathbb{R}, j \in [1, K]$: lower and upper bounds on the jth-descriptor;

variables:
- Non-negative integer variable $x(j) \in \mathbb{Z}_+, j \in I_{\mathbb{Z}} \cap I_+$;
- Integer variable $x(j) \in \mathbb{Z}, j \in I_{\mathbb{Z}} \setminus I_+$;
- Non-negative real variable $x(j) \in \mathbb{Z}_+, j \in I_+ \setminus I_{\mathbb{Z}}$;
- Real variable $x(j) \in \mathbb{Z}, j \in [1, K] \setminus (I_{\mathbb{Z}} \cup I_+)$;

constraints:

$$\ell(j) \le x(j) \le u(j), j \in [1, K], \tag{1}$$

$$\underline{y}^* \le \sum_{j\in[1,K]} w(j)x(j) + b \le \overline{y}^*, \tag{2}$$

objective function:
none.

The number of variables and constraints in the above MILP formulation is $O(K)$. It is not difficult to see that the above MILP is an NP-hard problem.

The entire MILP for Stage 4 consists of the two MILPs $\mathcal{M}(x, y; \mathcal{C}_1)$ and $\mathcal{M}(g, x; \mathcal{C}_2)$ with no objective function. The latter represents the computation process of our feature function f and a given topological specification. See Section 5 for the details of MILP $\mathcal{M}(g, x; \mathcal{C}_2)$.

2 A Full Description of Descriptors

Our definition of feature function is analogous with the one by Zhu et al. [1] except for a necessary modification due to our polymer model with link-edges.

Associated with the two functions α and β in a chemical graph $\mathbb{C} = (H, \alpha, \beta)$, we introduce functions $\mathrm{ac} : V(E) \to (\Lambda \setminus \{\mathtt{H}\}) \times (\Lambda \setminus \{\mathtt{H}\}) \times [1, 3]$, $\mathrm{cs} : V(E) \to (\Lambda \setminus \{\mathtt{H}\}) \times [1, 6]$ and $\mathrm{ec} : V(E) \to ((\Lambda \setminus \{\mathtt{H}\}) \times [1, 6]) \times ((\Lambda \setminus \{\mathtt{H}\}) \times [1, 6]) \times [1, 3]$ in the following.

To represent a feature of the exterior of $\mathbb{C}$, a chemical rooted tree in $\mathcal{T}(\mathbb{C})$ is called a *fringe-configuration* of $\mathbb{C}$.

We also represent leaf-edges in the exterior of $\mathbb{C}$. For a leaf-edge $uv \in E(\langle\mathbb{C}\rangle)$ with $\deg_{\langle\mathbb{C}\rangle}(u) = 1$, we define the *adjacency-configuration* of e to be an ordered tuple $(\alpha(u), \alpha(v), \beta(uv))$. Define

$$\Gamma^{\mathrm{lf}}_{\mathrm{ac}} \triangleq \{(\mathtt{a}, \mathtt{b}, m) \mid \mathtt{a}, \mathtt{b} \in \Lambda, m \in [1, \min\{\mathrm{val}(\mathtt{a}), \mathrm{val}(\mathtt{b})\}]\}$$

as a set of possible adjacency-configurations for leaf-edges.

To represent a feature of an interior-vertex $v \in V^{\mathrm{int}}(\mathbb{C})$ such that $\alpha(v) = \mathtt{a}$ and $\deg_{\langle\mathbb{C}\rangle}(v) = d$ (i.e., the number of non-hydrogen atoms adjacent to v is d) in a chemical graph $\mathbb{C} = (H, \alpha, \beta)$, we use a pair $(\mathtt{a}, d) \in (\Lambda \setminus \{\mathtt{H}\}) \times [1, 4]$, which we call the *chemical symbol* $\mathrm{cs}(v)$ of the vertex v. We treat $(\mathtt{a}, d)$ as a single symbol $\mathtt{a}d$, and define Λ_{dg} to be the set of all chemical symbols $\mu = \mathtt{a}d \in (\Lambda \setminus \{\mathtt{H}\}) \times [1, 4]$.

We define a method for featuring interior-edges as follows. Let $e = uv \in E^{\mathrm{int}}(\mathbb{C})$ be an interior-edge $e = uv \in E^{\mathrm{int}}(\mathbb{C})$ such that $\alpha(u) = \mathtt{a}$, $\alpha(v) = \mathtt{b}$ and $\beta(e) = m$ in a chemical graph $\mathbb{C} = (H, \alpha, \beta)$. To feature this edge e, we use a tuple $(\mathtt{a}, \mathtt{b}, m) \in (\Lambda \setminus \{\mathtt{H}\}) \times (\Lambda \setminus \{\mathtt{H}\}) \times [1, 3]$, which we call the *adjacency-configuration* $\mathrm{ac}(e)$ of the edge e. We introduce a total order $<$ over the elements in Λ to distinguish between $(\mathtt{a}, \mathtt{b}, m)$ and $(\mathtt{b}, \mathtt{a}, m)$ $(\mathtt{a} \neq \mathtt{b})$ notationally. For a tuple $\nu = (\mathtt{a}, \mathtt{b}, m)$, let $\overline{\nu}$ denote the tuple $(\mathtt{b}, \mathtt{a}, m)$.

Let $e = uv \in E^{\mathrm{int}}(\mathbb{C})$ be an interior-edge $e = uv \in E^{\mathrm{int}}(\mathbb{C})$ such that $\mathrm{cs}(u) = \mu$, $\mathrm{cs}(v) = \mu'$ and $\beta(e) = m$ in a chemical graph $\mathbb{C} = (H, \alpha, \beta)$. To feature this edge e, we use a tuple $(\mu, \mu', m) \in \Lambda_{\mathrm{dg}} \times \Lambda_{\mathrm{dg}} \times [1, 3]$, which we call the *edge-configuration* $\mathrm{ec}(e)$ of the edge e. We introduce a total order $<$ over the elements in Λ_{dg} to distinguish between (μ, μ', m) and (μ', μ, m) $(\mu \neq \mu')$ notationally. For a tuple $\gamma = (\mu, \mu', m)$, let $\overline{\gamma}$ denote the tuple (μ', μ, m).

Let π be a chemical property for which we will construct a prediction function η from a feature vector $f(\mathbb{C})$ of a chemical graph $\mathbb{C}$ to a predicted value $y \in \mathbb{R}$ for the chemical property of $\mathbb{C}$.

We first choose a set Λ of chemical elements and then collect a data set D_π of chemical compounds C whose chemical elements belong to Λ, where we regard D_π as a set of chemical graphs $\mathbb{C}$ that represent the chemical compounds C in D_π. To define the interior/exterior of chemical graphs $\mathbb{C} \in D_\pi$, we next choose a branch-parameter ρ, where we recommend $\rho = 2$.

Let $\Lambda^{\mathrm{int}}(D_\pi) \subseteq \Lambda$ (resp., $\Lambda^{\mathrm{ex}}(D_\pi) \subseteq \Lambda$) denote the set of chemical elements used in the set $V^{\mathrm{int}}(\mathbb{C})$ of interior-vertices (resp., the set $V^{\mathrm{ex}}(\mathbb{C})$ of exterior-vertices) of $\mathbb{C}$ over all chemical graphs $\mathbb{C} \in D_\pi$, and $\Gamma^{\mathrm{int}}(D_\pi)$ (resp., $\Gamma^{\mathrm{lnk}}(D_\pi)$) denote the set of edge-configurations used in the set $E^{\mathrm{int}}(\mathbb{C})$ of interior-edges (resp., the set $E^{\mathrm{lnk}}(\mathbb{C})$ of linked-edges) in $\mathbb{C}$ over all chemical graphs $\mathbb{C} \in D_\pi$. Let $\mathcal{F}(D_\pi)$ denote the set of chemical rooted trees ψ r-isomorphic to a chemical rooted tree in $\mathcal{T}(\mathbb{C})$ over all chemical graphs $\mathbb{C} \in D_\pi$, where possibly a chemical rooted tree $\psi \in \mathcal{F}(D_\pi)$ consists of a single chemical element $\mathtt{a} \in \Lambda \setminus \{\mathtt{H}\}$.

We define an integer 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. Introduce an integer coding of each of the sets $\Lambda^{\mathrm{int}}(D_\pi)$, $\Lambda^{\mathrm{ex}}(D_\pi)$, $\Gamma^{\mathrm{int}}(D_\pi)$ and $\mathcal{F}(D_\pi)$. Let $[\mathtt{a}]^{\mathrm{int}}$ (resp., $[\mathtt{a}]^{\mathrm{ex}}$) denote the coded integer of an element $\mathtt{a} \in \Lambda^{\mathrm{int}}(D_\pi)$ (resp., $\mathtt{a} \in \Lambda^{\mathrm{ex}}(D_\pi)$), $[\gamma]$ denote the coded integer of an element γ in $\Gamma^{\mathrm{int}}(D_\pi)$ and $[\psi]$ denote an element ψ in $\mathcal{F}(D_\pi)$.

We assume that a chemical graph $\mathbb{C}$ treated in this paper satisfies $\deg_{\langle\mathbb{C}\rangle}(v) \leq 4$ in the hydrogen-suppressed graph $\langle\mathbb{C}\rangle$.

In our model, we use an integer $\mathrm{mass}^*(\mathtt{a}) = \lfloor 10 \cdot \mathrm{mass}(\mathtt{a}) \rfloor$, for each $\mathtt{a} \in \Lambda$.

We define the *feature vector* $f(\mathbb{C})$ of a polymer $\mathbb{C} = (H, \alpha, \beta) \in D_\pi$ to be a vector that consists of the following non-negative integer descriptors $\mathrm{dcp}_i(\mathbb{C})$, $i \in [1, K]$, where $K = 14 + |\Lambda^{\mathrm{int}}(D_\pi)| + |\Lambda^{\mathrm{ex}}(D_\pi)| + |\Gamma^{\mathrm{int}}(D_\pi)| + |\Gamma^{\mathrm{lnk}}(D_\pi)| + |\Lambda_{\mathrm{dg}}| + |\mathcal{F}(D_\pi)| + |\Gamma^{\mathrm{lf}}_{\mathrm{ac}}|$.

1. $\mathrm{dcp}_1(\mathbb{C})$: the number $|V(H)| - |V_{\mathtt{H}}|$ of non-hydrogen atoms in $\mathbb{C}$.

2. $\mathrm{dcp}_2(\mathbb{C})$: the number $|V^{\mathrm{int}}(\mathbb{C})|$ of interior-vertices in $\mathbb{C}$.

3. $\mathrm{dcp}_3(\mathbb{C})$: the number $|E^{\mathrm{lnk}}(\mathbb{C})|$ of link-edges in $\mathbb{C}$. This descriptor is newly introduced in this paper to feature a structure of polymers.

4. $\mathrm{dcp}_4(\mathbb{C})$: the average $\overline{\mathrm{ms}}(\mathbb{C})$ of mass* over all atoms in $\mathbb{C}$;
   i.e., $\overline{\mathrm{ms}}(\mathbb{C}) \triangleq \frac{1}{|V(H)|}\sum_{v \in V(H)} \mathrm{mass}^*(\alpha(v))$.

5. $\mathrm{dcp}_i(\mathbb{C})$, $i = 4 + d, d \in [1, 4]$: the number $\mathrm{dg}^{\overline{\mathtt{H}}}_d(\mathbb{C})$ of non-hydrogen vertices $v \in V(H) \setminus V_{\mathtt{H}}$ of degree $\deg_{\langle\mathbb{C}\rangle}(v) = d$ in the hydrogen-suppressed chemical graph $\langle\mathbb{C}\rangle$.

6. $\mathrm{dcp}_i(\mathbb{C})$, $i = 8 + d, d \in [1, 4]$: the number $\mathrm{dg}^{\mathrm{int}}_d(\mathbb{C})$ of interior-vertices of interior-degree $\deg_{\mathbb{C}^{\mathrm{int}}}(v) = d$ in the interior $\mathbb{C}^{\mathrm{int}} = (V^{\mathrm{int}}(\mathbb{C}), E^{\mathrm{int}}(\mathbb{C}))$ of $\mathbb{C}$.

7. $\mathrm{dcp}_i(\mathbb{C})$, $i = 12 + m$, $m \in [2, 3]$: the number $\mathrm{bd}^{\mathrm{int}}_m(\mathbb{C})$ of interior-edges with bond multiplicity m in $\mathbb{C}$; i.e., $\mathrm{bd}^{\mathrm{int}}_m(\mathbb{C}) \triangleq \{e \in E^{\mathrm{int}}(\mathbb{C}) \mid \beta(e) = m\}$.

8. $\mathrm{dcp}_i(\mathbb{C})$, $i = 14 + [\mathtt{a}]^{\mathrm{int}}$, $\mathtt{a} \in \Lambda^{\mathrm{int}}(D_\pi)$: the frequency $\mathrm{na}^{\mathrm{int}}_{\mathtt{a}}(\mathbb{C}) = |V_{\mathtt{a}}(\mathbb{C}) \cap V^{\mathrm{int}}(\mathbb{C})|$ of chemical element $\mathtt{a}$ in the set $V^{\mathrm{int}}(\mathbb{C})$ of interior-vertices in $\mathbb{C}$.

9. $\mathrm{dcp}_i(\mathbb{C})$, $i = 14 + |\Lambda^{\mathrm{int}}(D_\pi)| + [\mathtt{a}]^{\mathrm{ex}}$, $\mathtt{a} \in \Lambda^{\mathrm{ex}}(D_\pi)$: the frequency $\mathrm{na}^{\mathrm{ex}}_{\mathtt{a}}(\mathbb{C}) = |V_{\mathtt{a}}(\mathbb{C}) \cap V^{\mathrm{ex}}(\mathbb{C})|$ of chemical element $\mathtt{a}$ in the set $V^{\mathrm{ex}}(\mathbb{C})$ of exterior-vertices in $\mathbb{C}$.

10. $\mathrm{dcp}_i(\mathbb{C})$, $i = 14 + |\Lambda^{\mathrm{int}}(D_\pi)| + |\Lambda^{\mathrm{ex}}(D_\pi)| + [\gamma]$, $\gamma \in \Gamma^{\mathrm{int}}(D_\pi)$: the frequency $\mathrm{ec}_\gamma(\mathbb{C})$ of edge-configuration γ in the set $E^{\mathrm{int}}(\mathbb{C})$ of interior-edges in $\mathbb{C}$.

11. $\mathrm{dcp}_i(\mathbb{C})$, $i = 14 + |\Lambda^{\mathrm{int}}(D_\pi)| + |\Lambda^{\mathrm{ex}}(D_\pi)| + |\Gamma^{\mathrm{int}}(D_\pi)| + [\gamma]$, $\gamma \in \Gamma^{\mathrm{lnk}}(D_\pi)$: the frequency $\mathrm{ec}_\gamma(\mathbb{C})$ of edge-configuration γ in the set $E^{\mathrm{lnk}}(\mathbb{C})$ of link-edges in $\mathbb{C}$. This descriptor is newly introduced in this paper to feature link-edges of polymers.

12. $\mathrm{dcp}_i(\mathbb{C})$, $i = 14 + |\Lambda^{\mathrm{int}}(D_\pi)| + |\Lambda^{\mathrm{ex}}(D_\pi)| + |\Gamma^{\mathrm{int}}(D_\pi)| + [\mu]$, $\mu \in \Lambda^{\mathrm{int}}_{\mathrm{dg}}$: the frequency of chemical symbols $\mu = \alpha(u)\deg_{\langle\mathbb{C}\rangle}(u)$ of connecting-vertices u in $\mathbb{C}$.

13. $\mathrm{dcp}_i(\mathbb{C})$, $i = 14 + |\Lambda^{\mathrm{int}}(D_\pi)| + |\Lambda^{\mathrm{ex}}(D_\pi)| + |\Gamma^{\mathrm{int}}(D_\pi)| + |\Gamma^{\mathrm{lnk}}(D_\pi)| + |\Lambda_{\mathrm{dg}}| + [\psi]$, $\psi \in \mathcal{F}(D_\pi)$: the frequency $\mathrm{fc}_\psi(\mathbb{C})$ of fringe-configuration ψ in the set of ρ-fringe-trees in $\mathbb{C}$.

14. $\mathrm{dcp}_i(\mathbb{C})$, $i = 14 + |\Lambda^{\mathrm{int}}(D_\pi)| + |\Lambda^{\mathrm{ex}}(D_\pi)| + |\Gamma^{\mathrm{int}}(D_\pi)| + |\Gamma^{\mathrm{lnk}}(D_\pi)| + |\Lambda_{\mathrm{dg}}| + |\mathcal{F}(D_\pi)| + [\nu]$, $\nu \in \Gamma^{\mathrm{lf}}_{\mathrm{ac}}$: the frequency $\mathrm{ac}^{\mathrm{lf}}_\nu(\mathbb{C})$ of adjacency-configuration ν in the set of leaf-edges in $\langle\mathbb{C}\rangle$.

3 Specifying Target Chemical Graphs

Our definition of topological specification is analogous with the one by Zhu et al. [1] except for a necessary modification due to our polymer model with link-edges.

Seed Graph

A *seed graph* for a polymer is defined to be a graph $G_{\mathrm{C}} = (V_{\mathrm{C}}, E_{\mathrm{C}})$ with a specified edge subset $E_{\mathrm{C}}^{\mathrm{lnk}}$ such that the edge set E_{C} consists of four sets $E_{(\geq 2)}$, $E_{(\geq 1)}$, $E_{(0/1)}$ and $E_{(=1)}$, where each of them can be empty, and $E_{\mathrm{C}}^{\mathrm{lnk}}$ is a circular set in G_{C} such that $\emptyset \neq E_{\mathrm{C}}^{\mathrm{lnk}} \subseteq E_{(\geq 2)} \cup E_{(\geq 1)} \cup E_{(=1)}$. Figure 5(a) illustrates an example of a seed graph, where $V_{\mathrm{C}} = \{u_1, u_2, \ldots, u_{14}\}$, $E_{(\geq 2)} = \{a_1, a_2, a_3, a_4\}$, $E_{(\geq 1)} = \{a_5, a_6, \ldots, a_9\}$, $E_{(0/1)} = \{a_{10}\}$, $E_{(=1)} = \{a_{11}, a_{12}, \ldots, a_{18}\}$ and $E_{\mathrm{C}}^{\mathrm{lnk}} = \{a_1, a_2\}$.

A *subdivision* S of G_{C} is a graph constructed from a seed graph G_{C} according to the following rules:

- Each edge $e = uv \in E_{(\geq 2)}$ is replaced with a u, v-path P_e of length at least 2;
- Each edge $e = uv \in E_{(\geq 1)}$ is replaced with a u, v-path P_e of length at least 1 (equivalently e is directly used or replaced with a u, v-path P_e of length at least 2);
- Each edge $e \in E_{(0/1)}$ is either used or discarded; and
- Each edge $e \in E_{(=1)}$ is always used directly.

The set of link-edges in the monomer representation $\mathbb{C}$ of an inferred polymer consists of edges in $E_{\mathrm{C}}^{\mathrm{lnk}} \cap (E_{(=1)} \cup E_{(\geq 1)})$ or edges in paths P_e for all edges $e = uv \in E_{\mathrm{C}}^{\mathrm{lnk}} \cap (E_{(\geq 1)} \cup E_{(\geq 2)})$ in a subdivision S of G_{C}.

A target chemical graph $\mathbb{C} = (H, \alpha, \beta)$ will contain S as a subgraph of the interior H^{int} of $\mathbb{C}$.

Interior-specification

A graph H^* that serves as the interior H^{int} of a target chemical graph $\mathbb{C}$ will be constructed as follows. First construct a subdivision S of a seed graph G_{C} by replacing each edge $e = uu' \in E_{(\geq 2)} \cup E_{(\geq 1)}$ with a pure u, u'-path P_e. Next construct a supergraph H^* of S by attaching a leaf path Q_v at each vertex $v \in V_{\mathrm{C}}$ or at an internal vertex $v \in V(P_e) \setminus \{u, u'\}$ of each pure u, u'-path P_e for some edge $e = uu' \in E_{(\geq 2)} \cup E_{(\geq 1)}$, where possibly $Q_v = (v), E(Q_v) = \emptyset$ (i.e., we do not attach any new edges to v). We introduce the following rules for specifying the size of H^*, the length $|E(P_e)|$ of a pure path P_e, the length $|E(Q_v)|$ of a leaf path Q_v, the number of leaf paths Q_v and a bond-multiplicity of each interior-edge, where we call the set of prescribed constants an *interior-specification* σ_{int}:

- Lower and upper bounds $\mathrm{n}_{\mathrm{LB}}^{\mathrm{int}}, \mathrm{n}_{\mathrm{UB}}^{\mathrm{int}} \in \mathbb{Z}_+$ on the number of interior-vertices of a target chemical graph $\mathbb{C}$.
- Lower and upper bounds $\mathrm{n}_{\mathrm{LB}}^{\mathrm{lnk}}, \mathrm{n}_{\mathrm{UB}}^{\mathrm{lnk}} \in \mathbb{Z}_+$ on the number of link-edges of a target chemical graph $\mathbb{C}$.
- For each edge $e = uu' \in E_{(\geq 2)} \cup E_{(\geq 1)}$,

  a lower bound $\ell_{\mathrm{LB}}(e)$ and an upper bound $\ell_{\mathrm{UB}}(e)$ on the length $|E(P_e)|$ of a pure u, u'-path P_e. (For a notational convenience, set $\ell_{\mathrm{LB}}(e) := 0$, $\ell_{\mathrm{UB}}(e) := 1$, $e \in E_{(0/1)}$ and $\ell_{\mathrm{LB}}(e) := 1$, $\ell_{\mathrm{UB}}(e) := 1$, $e \in E_{(=1)}$.)

  a lower bound $\mathrm{bl}_{\mathrm{LB}}(e)$ and an upper bound $\mathrm{bl}_{\mathrm{UB}}(e)$ on the number of leaf paths Q_v attached at internal vertices v of a pure u, u'-path P_e.

a lower bound $\mathrm{ch}_{\mathrm{LB}}(e)$ and an upper bound $\mathrm{ch}_{\mathrm{UB}}(e)$ on the maximum length $|E(Q_v)|$ of a leaf path Q_v attached at an internal vertex $v \in V(P_e) \setminus \{u, u'\}$ of a pure u, u'-path P_e.

- For each vertex $v \in V_{\mathrm{C}}$,

  a lower bound $\mathrm{ch}_{\mathrm{LB}}(v)$ and an upper bound $\mathrm{ch}_{\mathrm{UB}}(v)$ on the number of leaf paths Q_v attached to v, where $0 \leq \mathrm{ch}_{\mathrm{LB}}(v) \leq \mathrm{ch}_{\mathrm{UB}}(v) \leq 1$.

  a lower bound $\mathrm{ch}_{\mathrm{LB}}(v)$ and an upper bound $\mathrm{ch}_{\mathrm{UB}}(v)$ on the length $|E(Q_v)|$ of a leaf path Q_v attached to v.

- For each edge $e = uu' \in E_{\mathrm{C}}$, a lower bound $\mathrm{bd}_{m,\mathrm{LB}}(e)$ and an upper bound $\mathrm{bd}_{m,\mathrm{UB}}(e)$ on the number of edges with bond-multiplicity $m \in [2, 3]$ in u, u'-path P_e, where we regard P_e, $e \in E_{(0/1)} \cup E_{(=1)}$ as single edge e.

We call a graph H^* that satisfies an interior-specification σ_{int} a σ_{int}-*extension of* G_{C}, where the bond-multiplicity of each edge has been determined.

Table 1 shows an example of an interior-specification σ_{int} to the seed graph G_{C} in Figure 5(a).

Table 1: Example 1 of an interior-specification σ_{int}.

$\mathrm{n}^{\mathrm{int}}_{\mathrm{LB}} = 20$	$\mathrm{n}^{\mathrm{int}}_{\mathrm{UB}} = 30$	$\mathrm{n}^{\mathrm{lnk}}_{\mathrm{LB}} = 2$	$\mathrm{n}^{\mathrm{lnk}}_{\mathrm{UB}} = 24$

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
$\ell_{\mathrm{LB}}(a_i)$	2	4	3	2	2	1	1	1	1
$\ell_{\mathrm{UB}}(a_i)$	3	6	6	5	3	3	6	2	6
$\mathrm{bl}_{\mathrm{LB}}(a_i)$	0	1	1	0	0	0	0	0	0
$\mathrm{bl}_{\mathrm{UB}}(a_i)$	1	4	4	3	2	1	1	1	1
$\mathrm{ch}_{\mathrm{LB}}(a_i)$	0	2	1	0	0	0	0	0	0
$\mathrm{ch}_{\mathrm{UB}}(a_i)$	3	6	6	3	3	3	3	0	0

	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8	u_9	u_{10}	u_{11}	u_{12}	u_{13}	u_{14}
$\mathrm{bl}_{\mathrm{LB}}(u_i)$	0	0	0	0	0	0	0	0	1	0	0	0	0	0
$\mathrm{bl}_{\mathrm{UB}}(u_i)$	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\mathrm{ch}_{\mathrm{LB}}(u_i)$	0	0	0	0	0	0	0	0	1	0	0	0	0	0
$\mathrm{ch}_{\mathrm{UB}}(u_i)$	4	4	4	4	4	4	4	4	6	4	4	4	4	4

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}	a_{18}
$\mathrm{bd}_{2,\mathrm{LB}}(a_i)$	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
$\mathrm{bd}_{2,\mathrm{UB}}(a_i)$	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$\mathrm{bd}_{3,\mathrm{LB}}(a_i)$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\mathrm{bd}_{3,\mathrm{UB}}(a_i)$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Figure 6 illustrates an example of an σ_{int}-extension H^* of seed graph G_{C} in Figure 5(a) under the interior-specification σ_{int} in Table 1.

Chemical-specification

Let H^* be a graph that serves as the interior H^{int} of a target chemical graph $\mathbb{C}$, where the bond-multiplicity of each edge in H^* has be determined. Finally we introduce a set of rules for constructing a target chemical graph $\mathbb{C}$ from H^* by choosing a chemical element $\mathtt{a} \in \Lambda$ and assigning a ρ-fringe-tree ψ to each interior-vertex $v \in V^{\mathrm{int}}$. We introduce the following rules for specifying the size of $\mathbb{C}$, a set of chemical rooted trees that are allowed to use as ρ-fringe-trees and lower and upper bounds on the frequency of a chemical element, a chemical symbol, an edge-configuration, and a fringe-configuration where we call the set of prescribed constants a *chemical specification* σ_{ce}:

- Lower and upper bounds $n_{\mathrm{LB}}, n^* \in \mathbb{Z}_+$ on the number of vertices, where $\mathrm{n}^{\mathrm{int}}_{\mathrm{LB}} \leq n_{\mathrm{LB}} \leq n^*$.
- A subset $\mathcal{F}^* \subseteq \mathcal{F}(D_\pi)$ of chemical rooted trees ψ with $\mathrm{ht}(\langle\psi\rangle) \leq \rho$, where we require that every ρ-fringe-tree $\mathbb{C}[v]$ rooted at an interior-vertex v in $\mathbb{C}$ belongs to $\mathcal{F}^*$. Let Λ^{ex} denote the set of chemical elements assigned to non-root vertices over all chemical rooted trees in $\mathcal{F}^*$.
- A subset $\Lambda^{\mathrm{int}} \subseteq \Lambda^{\mathrm{int}}(D_\pi)$, where we require that every chemical element $\alpha(v)$ assigned to an interior-vertex v in $\mathbb{C}$ belongs to Λ^{int}. Let $\Lambda := \Lambda^{\mathrm{int}} \cup \Lambda^{\mathrm{ex}}$ and $\mathrm{na}_{\mathtt{a}}(\mathbb{C})$ (resp., $\mathrm{na}^{\mathrm{int}}_{\mathtt{a}}(\mathbb{C})$ and $\mathrm{na}^{\mathrm{ex}}_{\mathtt{a}}(\mathbb{C})$) denote the number of vertices (resp., interior-vertices and exterior-vertices) v such that $\alpha(v) = \mathtt{a}$ in $\mathbb{C}$.
- A set $\Lambda^{\mathrm{int}}_{\mathrm{dg}} \subseteq \Lambda \times [1, 4]$ of chemical symbols.
- Subsets $\Gamma^{\mathrm{lnk}} \subseteq \Gamma^{\mathrm{int}}$ of $\Gamma^{\mathrm{int}}(D_\pi)$ of edge-configurations (μ, μ', m) with $\mu \leq \mu'$, where we require that the edge-configuration $\mathrm{ec}(e)$ of an interior-edge (resp., a link-edge) e in $\mathbb{C}$ belongs to Γ^{int} (resp., Γ^{lnk}). We do not distinguish (μ, μ', m) and (μ', μ, m).
- Define $\Gamma^{\mathrm{int}}_{\mathrm{ac}}$ (resp., $\Gamma^{\mathrm{lnk}}_{\mathrm{ac}}$) to be the set of adjacency-configurations such that $\Gamma^{\mathrm{t}}_{\mathrm{ac}} := \{(\mathtt{a}, \mathtt{b}, m) \mid (\mathtt{a}d, \mathtt{b}d', m) \in \Gamma^{\mathrm{t}}\}, \mathrm{t} \in \{\mathrm{int}, \mathrm{lnk}\}$. Let $\mathrm{ac}^{\mathrm{int}}_\nu(\mathbb{C}), \nu \in \Gamma^{\mathrm{int}}_{\mathrm{ac}}$ (resp., $\mathrm{ac}^{\mathrm{lnk}}_\nu(\mathbb{C}), \nu \in \Gamma^{\mathrm{lnk}}_{\mathrm{ac}}$) denote the number of interior-edges (resp., link-edges) e such that $\mathrm{ac}(e) = \nu$ in $\mathbb{C}$.
- Subsets $\Lambda^*(v) \subseteq \{\mathtt{a} \in \Lambda^{\mathrm{int}} \mid \mathrm{val}(\mathtt{a}) \geq 2\}$, $v \in V_{\mathrm{C}}$, we require that every chemical element $\alpha(v)$ assigned to a vertex $v \in V_{\mathrm{C}}$ in the seed graph belongs to $\Lambda^*(v)$.
- Lower and upper bound functions $\mathrm{na}_{\mathrm{LB}}, \mathrm{na}_{\mathrm{UB}} : \Lambda \to [0, n^*]$ and $\mathrm{na}^{\mathrm{int}}_{\mathrm{LB}}, \mathrm{na}^{\mathrm{int}}_{\mathrm{UB}} : \Lambda^{\mathrm{int}} \to [0, n^*]$ on the number of interior-vertices v such that $\alpha(v) = \mathtt{a}$ in $\mathbb{C}$.
- Lower and upper bound functions $\mathrm{ns}^{\mathrm{int}}_{\mathrm{LB}}, \mathrm{ns}^{\mathrm{int}}_{\mathrm{UB}} : \Lambda^{\mathrm{int}}_{\mathrm{dg}} \to [0, n^*]$ on the number of interior-vertices v such that $\mathrm{cs}(v) = \mu$ in $\mathbb{C}$.
- Lower and upper bound functions $\mathrm{ns}^{\mathrm{cnt}}_{\mathrm{LB}}, \mathrm{ns}^{\mathrm{cnt}}_{\mathrm{UB}} : \Lambda^{\mathrm{int}}_{\mathrm{dg}} \to [0, 2]$ on the number of connecting-vertices v such that $\mathrm{cs}(v) = \mu$ in $\mathbb{C}$.
- Lower and upper bound functions $\mathrm{ac}^{\mathrm{int}}_{\mathrm{LB}}, \mathrm{ac}^{\mathrm{int}}_{\mathrm{UB}} : \Gamma^{\mathrm{int}}_{\mathrm{ac}} \to \mathbb{Z}_+$ ($\mathrm{ac}^{\mathrm{lnk}}_{\mathrm{LB}}, \mathrm{ac}^{\mathrm{lnk}}_{\mathrm{UB}} : \Gamma^{\mathrm{lnk}}_{\mathrm{ac}} \to \mathbb{Z}_+$) on the number of interior-edges (resp., link-edges) e such that $\mathrm{ac}(e) = \nu$ in $\mathbb{C}$.
- Lower and upper bound functions $\mathrm{ec}^{\mathrm{int}}_{\mathrm{LB}}, \mathrm{ec}^{\mathrm{int}}_{\mathrm{UB}} : \Gamma^{\mathrm{int}} \to \mathbb{Z}_+$ (resp., $\mathrm{ec}^{\mathrm{lnk}}_{\mathrm{LB}}, \mathrm{ec}^{\mathrm{lnk}}_{\mathrm{UB}} : \Gamma^{\mathrm{lnk}} \to \mathbb{Z}_+$) on the number of interior-edges (resp., link-edges) e such that $\mathrm{ec}(e) = \gamma$ in $\mathbb{C}$.
- Lower and upper bound functions $\mathrm{fc}_{\mathrm{LB}}, \mathrm{fc}_{\mathrm{UB}} : \mathcal{F}^* \to [0, n^*]$ on the number of interior-vertices v such that $\mathbb{C}[v]^{\mathrm{fr}}$ is r-isomorphic to $\psi \in \mathcal{F}^*$ in $\mathbb{C}$.
- Lower and upper bound functions $\mathrm{ac}^{\mathrm{lf}}_{\mathrm{LB}}, \mathrm{ac}^{\mathrm{lf}}_{\mathrm{UB}} : \Gamma^{\mathrm{lf}}_{\mathrm{ac}} \to [0, n^*]$ on the number of leaf-edges uv in ac_{C} with adjacency-configuration ν.

We call a chemical graph $\mathbb{C}$ that satisfies a chemical specification σ_{ce} a $(\sigma_{\mathrm{int}}, \sigma_{\mathrm{ce}})$-*extension of* G_{C}, and denote by $\mathcal{G}(G_{\mathrm{C}}, \sigma_{\mathrm{int}}, \sigma_{\mathrm{ce}})$ the set of all $(\sigma_{\mathrm{int}}, \sigma_{\mathrm{ce}})$-extensions of G_{C}.

Table 2 shows an example of a chemical-specification σ_{ce} to the seed graph G_{C} in Figure 5(a).

Figure 3 illustrates an example of a $(\sigma_{\mathrm{int}}, \sigma_{\mathrm{ce}})$-extension of G_{C} obtained from the σ_{int}-extension H^* in Figure 6 under the chemical-specification σ_{ce} in Table 2.

Table 2: Example 2 of a chemical-specification $\sigma_{\rm ce}$.

$n_{\rm LB} = 30$, $n^* = 50$.
branch-parameter: $\rho = 2$

Each of sets $\mathcal{F}(v), v \in V_{\rm C}$ and $\mathcal{F}_E$ is set to be the set $\mathcal{F}$ of chemical rooted trees ψ with $\mathrm{ht}(\langle\psi\rangle) \le \rho = 2$ in Figure 5(b).

$\Lambda = \{\mathtt{H}, \mathtt{C}, \mathtt{N}, \mathtt{O}, \mathtt{S}_{(2)}, \mathtt{S}_{(6)}, \mathtt{P} = \mathtt{P}_{(6)}, \mathtt{Cl}\}$	$\Lambda^{\rm int}_{\rm dg} = \{\mathtt{C2}, \mathtt{C3}, \mathtt{C4}, \mathtt{N2}, \mathtt{N3}, \mathtt{O2}, \mathtt{S}_{(2)}\mathtt{2}, \mathtt{S}_{(6)}\mathtt{3}, \mathtt{P4}\}$

$\Gamma^{\rm int}_{\rm ac}$	$\nu_1 = (\mathtt{C}, \mathtt{C}, 1), \nu_2 = (\mathtt{C}, \mathtt{C}, 2), \nu_3 = (\mathtt{C}, \mathtt{N}, 1), \nu_4 = (\mathtt{C}, \mathtt{O}, 1), \nu_5 = (\mathtt{C}, \mathtt{S}_{(2)}, 1), \nu_6 = (\mathtt{C}, \mathtt{S}_{(6)}, 1), \nu_7 = (\mathtt{C}, \mathtt{P}, 1)$
$\Gamma^{\rm int}$	$\gamma_1 = (\mathtt{C2}, \mathtt{C2}, 1), \gamma_2 = (\mathtt{C2}, \mathtt{C2}, 2), \gamma_3 = (\mathtt{C2}, \mathtt{C3}, 1), \gamma_4 = (\mathtt{C2}, \mathtt{C3}, 2), \gamma_5 = (\mathtt{C2}, \mathtt{C4}, 1), \gamma_6 = (\mathtt{C3}, \mathtt{C3}, 1),$ $\gamma_7 = (\mathtt{C3}, \mathtt{C3}, 2), \gamma_8 = (\mathtt{C3}, \mathtt{C4}, 1), \gamma_9 = (\mathtt{C2}, \mathtt{N3}, 1), \gamma_{10} = (\mathtt{C3}, \mathtt{N2}, 1), \gamma_{11} = (\mathtt{C4}, \mathtt{N2}, 1), \gamma_{12} = (\mathtt{C2}, \mathtt{O2}, 1),$ $\gamma_{13} = (\mathtt{C3}, \mathtt{O2}, 1), \gamma_{14} = (\mathtt{C2}, \mathtt{S}_{(2)}\mathtt{2}, 1), \gamma_{15} = (\mathtt{C3}, \mathtt{S}_{(2)}\mathtt{2}, 1), \gamma_{16} = (\mathtt{C4}, \mathtt{S}_{(2)}\mathtt{2}, 1), \gamma_{17} = (\mathtt{C3}, \mathtt{S}_{(6)}\mathtt{3}, 1),$ $\gamma_{18} = (\mathtt{C4}, \mathtt{S}_{(6)}\mathtt{3}, 1), \gamma_{19} = (\mathtt{C2}, \mathtt{P4}, 1), \gamma_{20} = (\mathtt{C3}, \mathtt{P4}, 1)$
$\Gamma^{\rm lnk}_{\rm ac}$	$\nu'_1 = (\mathtt{C}, \mathtt{C}, 1), \nu'_2 = (\mathtt{C}, \mathtt{C}, 2), \nu'_3 = (\mathtt{C}, \mathtt{N}, 1), \nu'_4 = (\mathtt{C}, \mathtt{S}_{(2)}, 1)$
$\Gamma^{\rm lnk}$	$\gamma'_1 = (\mathtt{C2}, \mathtt{C2}, 1), \gamma'_2 = (\mathtt{C2}, \mathtt{C3}, 1), \gamma'_3 = (\mathtt{C2}, \mathtt{C4}, 1), \gamma'_4 = (\mathtt{C3}, \mathtt{C3}, 1), \gamma'_5 = (\mathtt{C3}, \mathtt{C3}, 2), \gamma'_6 = (\mathtt{C2}, \mathtt{N3}, 1),$ $\gamma'_7 = (\mathtt{C3}, \mathtt{N2}, 1), \gamma'_8 = (\mathtt{C2}, \mathtt{S}_{(2)}\mathtt{2}, 1), \gamma'_9 = (\mathtt{C3}, \mathtt{S}_{(2)}\mathtt{2}, 1), \gamma'_{10} = (\mathtt{C4}, \mathtt{S}_{(2)}\mathtt{2}, 1)$

$\Lambda^*(u_i) = \{\mathtt{C}\}, i \in \{1, 2, 3, 4, 5, 6, 9\}$, $\Lambda^*(u_8) = \{\mathtt{O}\}$, $\Lambda^*(u_{12}) = \{\mathtt{C}, \mathtt{P}\}$, $\Lambda^*(u_i) = \{\mathtt{C}, \mathtt{O}, \mathtt{N}\}$, $i \in [1, 14] \setminus \{1, 2, 3, 4, 5, 6, 8, 9, 12\}$

	H	C	N	O	$\mathtt{S}_{(2)}$	$\mathtt{S}_{(6)}$	P	Cl
$\mathrm{na}_{\rm LB}(\mathtt{a})$	40	25	1	1	0	0	0	0
$\mathrm{na}_{\rm UB}(\mathtt{a})$	80	50	8	8	4	4	4	4

	C	N	O	$\mathtt{S}_{(2)}$	$\mathtt{S}_{(6)}$	P
$\mathrm{na}^{\rm int}_{\rm LB}(\mathtt{a})$	10	1	0	0	0	0
$\mathrm{na}^{\rm int}_{\rm UB}(\mathtt{a})$	25	4	5	2	2	2

	C2	C3	C4	N2	N3	O2	$\mathtt{S}_{(2)}\mathtt{2}$	$\mathtt{S}_{(6)}\mathtt{3}$	P4
$\mathrm{ns}^{\rm int}_{\rm LB}(\mu)$	3	5	0	0	0	0	0	0	0
$\mathrm{ns}^{\rm int}_{\rm UB}(\mu)$	12	15	5	5	3	5	1	1	1

	C2	C3	C4	N2	N3	O2	$\mathtt{S}_{(2)}\mathtt{2}$	$\mathtt{S}_{(6)}\mathtt{3}$	P4
$\mathrm{ns}^{\rm cnt}_{\rm LB}(\mu)$	0	0	0	0	0	0	0	0	0
$\mathrm{ns}^{\rm cnt}_{\rm UB}(\mu)$	2	2	2	2	2	2	1	1	0

	ν_1	ν_2	ν_3	ν_4	ν_5	ν_6	ν_7
$\mathrm{ac}^{\rm int}_{\rm LB}(\nu)$	0	0	0	0	0	0	0
$\mathrm{ac}^{\rm int}_{\rm UB}(\nu)$	30	10	10	10	2	3	3

	γ_1	γ_2	γ_3	γ_4	γ_5	$\gamma_i, i \in [6, 13]$	$\gamma_i, i \in [14, 20]$
$\mathrm{ec}^{\rm int}_{\rm LB}(\gamma)$	0	0	0	0	0	0	0
$\mathrm{ec}^{\rm int}_{\rm UB}(\gamma)$	4	15	5	5	10	5	2

	ν'_1	ν'_2	ν'_3	ν'_4
$\mathrm{ac}^{\rm lnk}_{\rm LB}(\nu')$	0	0	0	0
$\mathrm{ac}^{\rm lnk}_{\rm UB}(\nu')$	10	5	5	5

	$\gamma'_i, i \in [1, 10]$
$\mathrm{ec}^{\rm lnk}_{\rm LB}(\gamma')$	0
$\mathrm{ec}^{\rm lnk}_{\rm UB}(\gamma')$	4

	$\psi \in \{\psi_i \mid i = 1, 6, 11\}$	$\psi \in \mathcal{F}^* \setminus \{\psi_i \mid i = 1, 6, 11\}$
$\mathrm{fc}_{\rm LB}(\psi)$	1	0
$\mathrm{fc}_{\rm UB}(\psi)$	10	3

	$\nu \in \{(\mathtt{C}, \mathtt{C}, 1), (\mathtt{C}, \mathtt{C}, 2)\}$	$\nu \in \Gamma^{\rm lf}_{\rm ac} \setminus \{(\mathtt{C}, \mathtt{C}, 1), (\mathtt{C}, \mathtt{C}, 2)\}$
$\mathrm{ac}^{\rm lf}_{\rm LB}(\nu)$	0	0
$\mathrm{ac}^{\rm lf}_{\rm UB}(\nu)$	10	8

4 Test Instances for Stages 4 and 5

We prepared the following instances $I_{\rm a}$ and $I_{\rm b}$ for conducting experiments of Stages 4 and 5 in Phase 2.

In Stages 4 and 5, we use four properties $\pi \in \{$AmD, HcL, RfId, Tg$\}$ and define a set $\Lambda(\pi)$ of chemical elements as follows:
$\Lambda(\text{AmD}) = \Lambda_4 = \{\mathtt{H}, \mathtt{C}, \mathtt{N}, \mathtt{O}, \mathtt{Cl}, \mathtt{S}_{(2)}\}$, $\Lambda(\text{HcL}) = \Lambda(\text{Tg}) = \Lambda_5 = \{\mathtt{H}, \mathtt{C}, \mathtt{O}, \mathtt{N}, \mathtt{Cl}, \mathtt{S}_{(2)}, \mathtt{S}_{(6)}\}$, $\Lambda(\text{RfId}) = \Lambda_6 = \{\mathtt{H}, \mathtt{C}, \mathtt{O}_{(1)}, \mathtt{O}_{(2)}, \mathtt{N}, \mathtt{Cl}, \mathtt{Si}_{(4)}, \mathtt{F}\}$ and
$\Lambda(\text{Prm}) = \Lambda_3 = \{\mathtt{H}, \mathtt{C}, \mathtt{O}, \mathtt{N}, \mathtt{Cl}\}$.

(a) $I_{\rm a} = (G_{\rm C}, \sigma_{\rm int}, \sigma_{\rm ce})$: The instance used in Appendix 3 to explain the target specification. For each property $\pi \in \{$AmD, HcL, RfId, Tg, Prm$\}$, we replace $\Lambda = \{\mathtt{H}, \mathtt{C}, \mathtt{N}, \mathtt{O}, \mathtt{S}_{(2)}, \mathtt{S}_{(6)}, \mathtt{P}_{(5)}, \mathtt{Cl}\}$ in Table 2 with $\Lambda(\pi) \cap \{\mathtt{S}_{(2)}, \mathtt{S}_{(6)}, \mathtt{P}_{(5)}, \mathtt{Cl}\}$ and remove from the $\sigma_{\rm ce}$ all chemical symbols, edge-configurations and fringe-configurations that cannot be constructed from the replaced element set (i.e., those containing a chemical element in $\{\mathtt{S}_{(2)}, \mathtt{S}_{(6)}, \mathtt{P}_{(5)}, \mathtt{Cl}\} \setminus \Lambda(\pi)$).

(b) $I_{\rm b} = (G_{\rm C}, \sigma_{\rm int}, \sigma_{\rm ce})$: An instance that represents a set of polymers that includes the four examples of polymers in Figure 7. We set a seed graph $G_{\rm C} = (V_{\rm C}, E_{\rm C} = E_{(=1)})$ to be the graph with two cycles C_1 and C_2 in Figure 8(i), where we set $E_{(\geq 2)} = E_{\rm C}^{\rm lnk} = \{a_1, a_2\}$ and $E_{(=1)} = \{a_3, a_{12}, \ldots, a_{14}\}$.
Set $\Lambda := \Lambda(\pi)$ for each property $\pi \in \{$AmD, HcL, RfId, Tg$\}$, and set $\Lambda_{\rm dg}^{\rm int}$ to be the set of all possible chemical symbols in $\Lambda \times [1, 4]$.
Set $\Gamma^{\rm int}$ (resp., $\Gamma^{\rm lnk}$) to be the set of edge-configurations of the interior-edges (resp., the link-edges) used in the four examples of polymers in Figure 7. Set $\Gamma_{\rm ac}^{\rm int}$ (resp., $\Gamma_{\rm ac}^{\rm lnk}$) to be the set of the adjacency-configurations of the edge-configurations in $\Gamma^{\rm int}$ (resp., $\Gamma^{\rm lnk}$).
We specify $n_{\rm LB}$ for each property π and set $\mathrm{n}_{\rm LB}^{\rm int} := 14$, $\mathrm{n}_{\rm UB}^{\rm int} := n^* := n_{\rm LB} + 10$, $\mathrm{n}_{\rm LB}^{\rm lnk} := 2$, $\mathrm{n}_{\rm UB}^{\rm lnk} := 2 + \max\{n_{\rm LB} - 15, 0\}$.
For each link-edge $a_i \in E_{(\geq 2)} = E_{\rm C}^{\rm lnk} = \{a_1, a_2\}$, set $\ell_{\rm LB}(a_i) := 2 + \max\{\lfloor (n_{\rm LB} - 15)/4 \rfloor, 0\}$, $\ell_{\rm UB}(a_i) := \ell_{\rm LB}(a_i) + 5$, $\mathrm{bl}_{\rm LB}(a_i) := 0, \mathrm{bl}_{\rm UB}(a_i) := 3$, $\mathrm{ch}_{\rm LB}(a_i) := 0, \mathrm{ch}_{\rm UB}(a_i) := 5$, $\mathrm{bd}_{2,\rm LB}(a_i) := 0$ and $\mathrm{bd}_{2,\rm UB}(a_i) := \lfloor \ell_{\rm LB}(a_i)/3 \rfloor$.
To form two benzene rings from the two cycles C_1 and C_2, set $\Lambda^*(u) := \{\mathtt{C}\}$, $\mathrm{bl}_{\rm LB}(u) := \mathrm{bl}_{\rm UB}(u) := \mathrm{ch}_{\rm LB}(u) := \mathrm{ch}_{\rm UB}(u) := 0, u \in V_{\rm C}$, $\mathrm{bd}_{2,\rm LB}(a_i) := \mathrm{bd}_{2,\rm UB}(a_i) := 0, i \in \{3, 5, 7, 9, 11, 13\}$, $\mathrm{bd}_{2,\rm LB}(a_i) := \mathrm{bd}_{2,\rm UB}(a_i) := 1, i \in \{4, 6, 8, 10, 12, 14\}$.
Not to include any triple-bond, set $\mathrm{bd}_{3,\rm LB}(a) := \mathrm{bd}_{3,\rm UB}(a) := 0, a \in E_{\rm C}$.
Set lower bounds $\mathrm{na}_{\rm LB}$, $\mathrm{na}_{\rm LB}^{\rm int}$, $\mathrm{ns}_{\rm LB}^{\rm int}$, $\mathrm{ns}_{\rm LB}^{\rm cnt}$, $\mathrm{ac}_{\rm LB}^{\rm int}$, $\mathrm{ac}_{\rm LB}^{\rm lnk}$, $\mathrm{ec}_{\rm LB}^{\rm int}$, $\mathrm{ec}_{\rm LB}^{\rm lnk}$ and $\mathrm{ac}_{\rm LB}^{\rm lf}$ to be 0.
Set upper bounds $\mathrm{na}_{\rm UB}(\mathtt{a}) := n^*, \mathrm{na} \in \{\mathtt{H}, \mathtt{C}\}$, $\mathrm{na}_{\rm UB}(\mathtt{a}) := 5 + \max\{n_{\rm LB} - 15, 0\}, \mathtt{a} \in \{\mathtt{O}, \mathtt{N}\}$, $\mathrm{na}_{\rm UB}(\mathtt{a}) := 2 + \max\{\lfloor (n_{\rm LB} - 15)/4 \rfloor, 0\}, \mathtt{a} \in \Lambda \setminus \{\mathtt{H}, \mathtt{C}, \mathtt{O}, \mathtt{N}\}$, $\mathrm{ns}_{\rm UB}^{\rm cnt}(\mu) := 2, \mu \in \Lambda_{\rm dg}^{\rm int}$, and $\mathrm{na}_{\rm UB}^{\rm int}$, $\mathrm{ns}_{\rm UB}^{\rm int}$, $\mathrm{ac}_{\rm UB}^{\rm int}$, $\mathrm{ac}_{\rm LB}^{\rm lnk}$, $\mathrm{ec}_{\rm UB}^{\rm int}$, $\mathrm{ec}_{\rm UB}^{\rm lnk}$ and $\mathrm{ac}_{\rm UB}^{\rm lf}$ to be n^*.
Set $\mathcal{F}$ to be the set of the 17 chemical rooted trees $\psi_i, i \in [1, 17]$ in Figure 8(ii). Set $\mathcal{F}_E := \mathcal{F}(v) := \mathcal{F}$, $v \in V_{\rm C}$ and $\mathrm{fc}_{\rm LB}(\psi) := 0, \psi \in \mathcal{F}$, $\mathrm{fc}_{\rm UB}(\psi_i) := 12 + \max\{n_{\rm LB} - 15, 0\}, i \in [1, 4]$, $\mathrm{fc}_{\rm UB}(\psi_i) := 8 + \max\{\lfloor (n_{\rm LB} - 15)/2 \rfloor, 0\}, i \in [5, 12]$ and $\mathrm{fc}_{\rm UB}(\psi_i) := 5 + \max\{\lfloor (n_{\rm LB} - 15)/4 \rfloor, 0\}, i \in [13, 17], \psi_i \in \mathcal{F}$.

5 All Constraints in an MILP Formulation for Chemical Graphs

Our definition of an MILP formulation MILP $\mathcal{M}(g, x; \mathcal{C}_2)$ is analogous with the one by Zhu et al. [1] except for a necessary modification due to our polymer model with link-edges.

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 $[\mathtt{e}]$ the encoded element $\sigma(\mathtt{e})$. Let ϵ denote *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_ϵ denote the set $A \cup \{\epsilon\}$ and define a standard encoding of A_ϵ to be a bijection $\sigma : A \to [0, |A|]$ such that $\sigma(\epsilon) = 0$, where we denote by $[A_\epsilon]$ the set $[0, |A|]$ of integers and by $[\mathtt{e}]$ the encoded element $\sigma(\mathtt{e})$, where $[\epsilon] = 0$.

Let $\sigma = (G_{\mathrm{C}}, \sigma_{\mathrm{int}}, \sigma_{\mathrm{ce}})$ be a target specification, ρ denote the branch-parameter in the specification σ and $\mathbb{C}$ denote a chemical graph in $\mathcal{G}(G_{\mathrm{C}}, \sigma_{\mathrm{int}}, \sigma_{\mathrm{ce}})$.

5.1 Selecting a Cyclical-base

Recall that

$$\begin{aligned} &E_{(=1)} = \{e \in E_{\mathrm{C}} \mid \ell_{\mathrm{LB}}(e) = \ell_{\mathrm{UB}}(e) = 1\}; && E_{(0/1)} = \{e \in E_{\mathrm{C}} \mid \ell_{\mathrm{LB}}(e) = 0, \ell_{\mathrm{UB}}(e) = 1\}; \\ &E_{(\geq 1)} = \{e \in E_{\mathrm{C}} \mid \ell_{\mathrm{LB}}(e) = 1, \ell_{\mathrm{UB}}(e) \geq 2\}; && E_{(\geq 2)} = \{e \in E_{\mathrm{C}} \mid \ell_{\mathrm{LB}}(e) \geq 2\}; \end{aligned}$$

A subset $E_{\mathrm{C}}^{\mathrm{lnk}} \subseteq E_{(=1)} \cup E_{(\geq 1)} \cup E_{(\geq 2)}$ is given for introducing link-edges in the monomer representation $\mathbb{C}$ of an inferred polymer.

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

  $$P_i = ({v^{\mathrm{C}}}_{\mathrm{tail}(i)}, {v^{\mathrm{T}}}_{j-1}, {v^{\mathrm{T}}}_{j}, \ldots, {v^{\mathrm{T}}}_{j+t}, {v^{\mathrm{C}}}_{\mathrm{head}(i)})$$

  of length at least 2 from vertex ${v^{\mathrm{C}}}_{\mathrm{tail}(i)}$ to vertex ${v^{\mathrm{C}}}_{\mathrm{head}(i)}$ visiting some vertices in V_{T} is constructed in $\langle \mathbb{C} \rangle$; and
- For each edge $a_i \in E_{(\geq 1)}$, either edge a_i is directly used in $\langle \mathbb{C} \rangle$ or the above path P_i of length at least 2 is constructed in $\langle \mathbb{C} \rangle$.

Let $t_{\mathrm{C}} \triangleq |V_{\mathrm{C}}|$ and denote V_{C} by $\{{v^{\mathrm{C}}}_i \mid i \in [1, t_{\mathrm{C}}]\}$. Regard the seed graph G_{C} as a digraph such that each edge a_i with end-vertices ${v^{\mathrm{C}}}_j$ and ${v^{\mathrm{C}}}_{j'}$ is directed from ${v^{\mathrm{C}}}_j$ to ${v^{\mathrm{C}}}_{j'}$ when $j < j'$. For each directed edge $a_i \in E_{\mathrm{C}}$, let head(i) and tail(i) denote the head and tail of $e^{\mathrm{C}}(i)$; i.e., $a_i = ({v^{\mathrm{C}}}_{\mathrm{tail}(i)}, {v^{\mathrm{C}}}_{\mathrm{head}(i)})$.

Define

$$k_{\mathrm{C}} \triangleq |E_{(\geq 2)} \cup E_{(\geq 1)}|, \quad \widetilde{k_{\mathrm{C}}} \triangleq |E_{(\geq 2)}|,$$

and denote $E_{\mathrm{C}} = \{a_i \mid i \in [1, m_{\mathrm{C}}]\}$,

$$E_{(\geq 2)} = \{a_k \mid k \in [1, \widetilde{k_{\mathrm{C}}}]\}, \; E_{(\geq 1)} = \{a_k \mid k \in [\widetilde{k_{\mathrm{C}}} + 1, k_{\mathrm{C}}]\},$$

$$E_{(0/1)} = \{a_i \mid i \in [k_{\mathrm{C}} + 1, k_{\mathrm{C}} + |E_{(0/1)}|]\} \text{ and } E_{(=1)} = \{a_i \mid i \in [k_{\mathrm{C}} + |E_{(0/1)}| + 1, m_{\mathrm{C}}]\}.$$

Let $I_{(=1)}$ denote the set of indices i of edges $a_i \in E_{(=1)}$. Similarly for $I_{(0/1)}$, $I_{(\geq 1)}$ and $I_{(\geq 2)}$. Let $I_{\rm lnk}$ denote the set of indices i of edges $a_i \in E_{\rm C}^{\rm lnk}$.

To control the construction of such a path P_i for each edge $a_k \in E_{(\geq 2)} \cup E_{(\geq 1)}$, we regard the index $k \in [1, k_{\rm C}]$ of each edge $a_k \in E_{(\geq 2)} \cup E_{(\geq 1)}$ 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 $v^{\rm T}{}_{j-1}, v^{\rm T}{}_j, \ldots, v^{\rm T}{}_{j+t}$ in $V_{\rm T}$ when the above path P_k is used in $\langle \mathbb{C} \rangle$.

For each index $s \in [1, t_{\rm C}]$, let $I_{\rm C}(s)$ denote the set of edges $e \in E_{\rm C}$ incident to vertex $v^{\rm C}{}_s$, 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 $v^{\rm C}{}_s$. Similarly for $E^+_{(0/1)}(s)$, $E^-_{(0/1)}(s)$, $E^+_{(\geq 1)}(s)$, $E^-_{(\geq 1)}(s)$, $E^+_{(\geq 2)}(s)$ and $E^-_{(\geq 2)}(s)$. Let $I_{\rm C}(s)$ denote the set of indices i of edges $a_i \in I_{\rm C}(s)$. Similarly for $I^+_{(=1)}(s)$, $I^-_{(=1)}(s)$, $I^+_{(0/1)}(s)$, $I^-_{(0/1)}(s)$, $I^+_{(\geq 1)}(s)$, $I^-_{(\geq 1)}(s)$, $I^+_{(\geq 2)}(s)$ and $I^-_{(\geq 2)}(s)$. Note that $[1, k_{\rm C}] = I_{(\geq 2)} \cup I_{(\geq 1)}$ and $[\widetilde{k_{\rm C}}+1, m_{\rm C}] = I_{(\geq 1)} \cup I_{(0/1)} \cup I_{(=1)}$.

constants:

- $n^* \in \mathbb{Z}$: an upper bound on the number $n(\mathbb{C})$ of non-hydrogen atoms in $\mathbb{C}$;
- $t_{\rm C} = |V_{\rm C}|$, $\widetilde{k_{\rm C}} = |E_{(\geq 2)}|$, $k_{\rm C} = |E_{(\geq 2)} \cup E_{(\geq 1)}|$, $t_{\rm T} = {\rm n}^{\rm int}_{\rm UB} - |V_{\rm C}|$, $m_{\rm C} = |E_{\rm C}|$. Note that $a_i \in E_{\rm C} \setminus (E_{(\geq 2)} \cup E_{(\geq 1)})$ holds $i \in [k_{\rm C}+1, m_{\rm C}]$;
- $\ell_{\rm LB}(k), \ell_{\rm UB}(k) \in [1, t_{\rm T}]$, $k \in [1, k_{\rm C}]$: lower and upper bounds on the length of path P_k;
- $n^{(=1)}_{\rm lnk} = |I_{\rm lnk} \cap E_{(=1)}| = |I_{\rm lnk} \cap \{[k_{\rm C} + |E_{(0/1)}| + 1, m_{\rm C}\}|$: the number of link-edges created from $E_{(=1)}$;
- ${\rm n}^{\rm lnk}_{\rm LB}, {\rm n}^{\rm lnk}_{\rm UB} \in [0, n^*]$: lower and upper bounds on the number of link-edges of a target polymer $\mathbb{C}$;

variables:

- $e^{\rm C}(i) \in [0,1]$, $i \in [1, m_{\rm C}]$: $e^{\rm C}(i)$ represents edge $a_i \in E_{\rm C}$, $i \in [1, m_{\rm C}]$ ($e^{\rm C}(i) = 1$, $i \in I_{(=1)}$; $e^{\rm C}(i) = 0$, $i \in I_{(\geq 2)}$) ($e^{\rm C}(i) = 1 \Leftrightarrow$ edge a_i is used in $\langle \mathbb{C} \rangle$);
- $v^{\rm T}(i) \in [0,1]$, $i \in [1, t_{\rm T}]$: $v^{\rm T}(i) = 1 \Leftrightarrow$ vertex $v^{\rm T}{}_i$ is used in $\langle \mathbb{C} \rangle$;
- $e^{\rm T}(i) \in [0,1]$, $i \in [1, t_{\rm T}+1]$: $e^{\rm T}(i)$ represents edge $e^{\rm T}{}_i = (v^{\rm T}{}_{i-1}, v^{\rm T}{}_i) \in E_{\rm T}$, where $e^{\rm T}{}_1$ and $e^{\rm T}{}_{t_{\rm T}+1}$ are fictitious edges ($e^{\rm T}(i) = 1 \Leftrightarrow$ edge $e^{\rm T}{}_i$ is used in $\langle \mathbb{C} \rangle$);
- $\chi^{\rm T}(i) \in [0, k_{\rm C}]$, $i \in [1, t_{\rm T}]$: $\chi^{\rm T}(i)$ represents the color assigned to vertex $v^{\rm T}{}_i$ ($\chi^{\rm T}(i) = k > 0 \Leftrightarrow$ vertex $v^{\rm T}{}_i$ is assigned color k; $\chi^{\rm T}(i) = 0$ means that vertex $v^{\rm T}{}_i$ is not used in $\langle \mathbb{C} \rangle$);
- ${\rm clr}^{\rm T}(k) \in [\ell_{\rm LB}(k) - 1, \ell_{\rm UB}(k) - 1]$, $k \in [1, k_{\rm C}]$, ${\rm clr}^{\rm T}(0) \in [0, t_{\rm T}]$: the number of vertices $v^{\rm T}{}_i \in V_{\rm T}$ with color c;
- $\delta^{\rm T}_\chi(k) \in [0,1]$, $k \in [0, k_{\rm C}]$: $\delta^{\rm T}_\chi(k) = 1 \Leftrightarrow \chi^{\rm T}(i) = k$ for some $i \in [1, t_{\rm T}]$;
- $\chi^{\rm T}(i,k) \in [0,1]$, $i \in [1, t_{\rm T}]$, $k \in [0, k_{\rm C}]$ ($\chi^{\rm T}(i,k) = 1 \Leftrightarrow \chi^{\rm T}(i) = k$);
- $\widetilde{\rm deg}^+_{\rm C}(i) \in [0,4]$, $i \in [1, t_{\rm C}]$: the out-degree of vertex $v^{\rm C}{}_i$ with the used edges $e^{\rm C}$ in $E_{\rm C}$;
- $\widetilde{\rm deg}^-_{\rm C}(i) \in [0,4]$, $i \in [1, t_{\rm C}]$: the in-degree of vertex $v^{\rm C}{}_i$ with the used edges $e^{\rm C}$ in $E_{\rm C}$;
- $n_{\rm lnk} \in [{\rm n}^{\rm lnk}_{\rm LB}, {\rm n}^{\rm lnk}_{\rm UB}]$: the number of link-edges in $\mathbb{C}$;

constraints:

$$e^{\rm C}(i) = 1, \quad i \in I_{(=1)}, \tag{3}$$

$$e^{\rm C}(i) = 0, \ \ {\rm clr}^{\rm T}(i) \geq 1, \quad i \in I_{(\geq 2)}, \tag{4}$$

$$e^{\rm C}(i) + {\rm clr}^{\rm T}(i) \geq 1, \qquad {\rm clr}^{\rm T}(i) \leq t_{\rm T} \cdot (1 - e^{\rm C}(i)), \quad i \in I_{(\geq 1)}, \tag{5}$$

$$\sum_{c \in I^{-}_{(\geq 1)}(i) \cup I^{-}_{(0/1)}(i) \cup I^{-}_{(=1)}(i)} e^{\rm C}(c) = \widetilde{\deg}^{-}_{\rm C}(i), \qquad \sum_{c \in I^{+}_{(\geq 1)}(i) \cup I^{+}_{(0/1)}(i) \cup I^{+}_{(=1)}(i)} e^{\rm C}(c) = \widetilde{\deg}^{+}_{\rm C}(i), \quad i \in [1, t_{\rm C}], \tag{6}$$

$$\chi^{\rm T}(i, 0) = 1 - v^{\rm T}(i), \qquad \sum_{k \in [0, k_{\rm C}]} \chi^{\rm T}(i, k) = 1, \qquad \sum_{k \in [0, k_{\rm C}]} k \cdot \chi^{\rm T}(i, k) = \chi^{\rm T}(i), \qquad i \in [1, t_{\rm T}], \tag{7}$$

$$\sum_{i \in [1, t_{\rm T}]} \chi^{\rm T}(i, k) = {\rm clr}^{\rm T}(k), \quad t_{\rm T} \cdot \delta^{\rm T}_{\chi}(k) \geq \sum_{i \in [1, t_{\rm T}]} \chi^{\rm T}(i, k) \geq \delta^{\rm T}_{\chi}(k), \qquad k \in [0, k_{\rm C}], \tag{8}$$

$$\begin{gathered} v^{\rm T}(i-1) \geq v^{\rm T}(i), \\ k_{\rm C} \cdot (v^{\rm T}(i-1) - e^{\rm T}(i)) \geq \chi^{\rm T}(i-1) - \chi^{\rm T}(i) \geq v^{\rm T}(i-1) - e^{\rm T}(i), \end{gathered} \qquad i \in [2, t_{\rm T}], \tag{9}$$

$$\sum_{k \in I_{\rm lnk} \cap [1, k_{\rm C}]} ({\rm clr}^{\rm T}(k) + 1) + n^{(=1)}_{\rm lnk} = n_{\rm lnk}. \tag{10}$$

5.2 Constraints for Including Leaf Paths

Let $\widetilde{t_{\rm C}}$ denote the number of vertices $u \in V_{\rm C}$ such that ${\rm bl}_{\rm UB}(u) = 1$ and assume that $V_{\rm C} = \{u_1, u_2, \ldots, u_p\}$ so that

$${\rm bl}_{\rm UB}(u_i) = 1,\ i \in [1, \widetilde{t_{\rm C}}] \text{ and } {\rm bl}_{\rm UB}(u_i) = 0,\ i \in [\widetilde{t_{\rm C}} + 1, t_{\rm C}].$$

Define the set of colors for the vertex set $\{u_i \mid i \in [1, \widetilde{t_{\rm C}}]\} \cup V_{\rm T}$ to be $[1, c_{\rm F}]$ with

$$c_{\rm F} \triangleq \widetilde{t_{\rm C}} + t_{\rm T} = |\{u_i \mid i \in [1, \widetilde{t_{\rm C}}]\} \cup V_{\rm T}|.$$

Let each vertex ${v^{\rm C}}_i$, $i \in [1, \widetilde{t_{\rm C}}]$ (resp., ${v^{\rm T}}_i \in V_{\rm T}$) correspond to a color $i \in [1, c_{\rm F}]$ (resp., $i + \widetilde{t_{\rm C}} \in [1, c_{\rm F}]$). When a path $P = (u, {v^{\rm F}}_j, {v^{\rm F}}_{j+1}, \ldots, {v^{\rm F}}_{j+t})$ from a vertex $u \in V_{\rm C} \cup V_{\rm T}$ is used in $\langle \mathbb{C} \rangle$, we assign the color $i \in [1, c_{\rm F}]$ of the vertex u to the vertices ${v^{\rm F}}_j, {v^{\rm F}}_{j+1}, \ldots, {v^{\rm F}}_{j+t} \in V_{\rm F}$.

constants:

- $c_{\rm F}$: the maximum number of different colors assigned to the vertices in $V_{\rm F}$;
- ${\rm n}^{\rm int}_{\rm LB}, {\rm n}^{\rm int}_{\rm UB} \in [2, n^*]$: lower and upper bounds on the number of interior-vertices in $\mathbb{C}$;
- ${\rm bl}_{\rm LB}(i) \in [0, 1]$, $i \in [1, \widetilde{t_{\rm C}}]$: a lower bound on the number of leaf ρ-branches in the leaf path rooted at a vertex ${v^{\rm C}}_i$;

- $\mathrm{bl}_{\mathrm{LB}}(k), \mathrm{bl}_{\mathrm{UB}}(k) \in [0, \ell_{\mathrm{UB}}(k)-1]$, $k \in [1, k_{\mathrm{C}}] = I_{(\geq 2)} \cup I_{(\geq 1)}$: lower and upper bounds on the number of leaf ρ-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:

- $\mathrm{n}_G^{\mathrm{int}} \in [\mathrm{n}_{\mathrm{LB}}^{\mathrm{int}}, \mathrm{n}_{\mathrm{UB}}^{\mathrm{int}}]$: the number of interior-vertices in $\mathbb{C}$;
- $v^{\mathrm{F}}(i) \in [0,1]$, $i \in [1, t_{\mathrm{F}}]$: $v^{\mathrm{F}}(i) = 1 \Leftrightarrow$ vertex ${v^{\mathrm{F}}}_i$ is used in $\mathbb{C}$;
- $e^{\mathrm{F}}(i) \in [0,1]$, $i \in [1, t_{\mathrm{F}}+1]$: $e^{\mathrm{F}}(i)$ represents edge ${e^{\mathrm{F}}}_i = {v^{\mathrm{F}}}_{i-1}{v^{\mathrm{F}}}_i$, where ${e^{\mathrm{F}}}_1$ and ${e^{\mathrm{F}}}_{t_{\mathrm{F}}+1}$ are fictitious edges ($e^{\mathrm{F}}(i) = 1 \Leftrightarrow$ edge ${e^{\mathrm{F}}}_i$ is used in $\mathbb{C}$);
- $\chi^{\mathrm{F}}(i) \in [0, c_{\mathrm{F}}]$, $i \in [1, t_{\mathrm{F}}]$: $\chi^{\mathrm{F}}(i)$ represents the color assigned to vertex ${v^{\mathrm{F}}}_i$ ($\chi^{\mathrm{F}}(i) = c \Leftrightarrow$ vertex ${v^{\mathrm{F}}}_i$ is assigned color c);
- $\mathrm{clr}^{\mathrm{F}}(c) \in [0, t_{\mathrm{F}}]$, $c \in [0, c_{\mathrm{F}}]$: the number of vertices ${v^{\mathrm{F}}}_i$ with color c;
- $\delta_\chi^{\mathrm{F}}(c) \in [\mathrm{bl}_{\mathrm{LB}}(c), 1]$, $c \in [1, \widetilde{t_{\mathrm{C}}}]$: $\delta_\chi^{\mathrm{F}}(c) = 1 \Leftrightarrow \chi^{\mathrm{F}}(i) = c$ for some $i \in [1, t_{\mathrm{F}}]$;
- $\delta_\chi^{\mathrm{F}}(c) \in [0, 1]$, $c \in [\widetilde{t_{\mathrm{C}}}+1, c_{\mathrm{F}}]$: $\delta_\chi^{\mathrm{F}}(c) = 1 \Leftrightarrow \chi^{\mathrm{F}}(i) = c$ for some $i \in [1, t_{\mathrm{F}}]$;
- $\chi^{\mathrm{F}}(i,c) \in [0,1]$, $i \in [1, t_{\mathrm{F}}]$, $c \in [0, c_{\mathrm{F}}]$: $\chi^{\mathrm{F}}(i,c) = 1 \Leftrightarrow \chi^{\mathrm{F}}(i) = c$;
- $\mathrm{bl}(k,i) \in [0,1]$, $k \in [1, k_{\mathrm{C}}] = I_{(\geq 2)} \cup I_{(\geq 1)}$, $i \in [1, t_{\mathrm{T}}]$: $\mathrm{bl}(k,i) = 1 \Leftrightarrow$ path P_k contains vertex ${v^{\mathrm{T}}}_i$ as an internal vertex and the ρ-fringe-tree rooted at ${v^{\mathrm{T}}}_i$ contains a leaf ρ-branch;

constraints:

$$\chi^{\mathrm{F}}(i,0) = 1 - v^{\mathrm{F}}(i), \qquad \sum_{c \in [0, c_{\mathrm{F}}]} \chi^{\mathrm{F}}(i,c) = 1, \qquad \sum_{c \in [0, c_{\mathrm{F}}]} c \cdot \chi^{\mathrm{F}}(i,c) = \chi^{\mathrm{F}}(i), \qquad i \in [1, t_{\mathrm{F}}], \tag{11}$$

$$\sum_{i \in [1, t_{\mathrm{F}}]} \chi^{\mathrm{F}}(i,c) = \mathrm{clr}^{\mathrm{F}}(c), \quad t_{\mathrm{F}} \cdot \delta_\chi^{\mathrm{F}}(c) \geq \sum_{i \in [1, t_{\mathrm{F}}]} \chi^{\mathrm{F}}(i,c) \geq \delta_\chi^{\mathrm{F}}(c), \qquad c \in [0, c_{\mathrm{F}}], \tag{12}$$

$$e^{\mathrm{F}}(1) = e^{\mathrm{F}}(t_{\mathrm{F}}+1) = 0, \tag{13}$$

$$\begin{aligned} v^{\mathrm{F}}(i-1) &\geq v^{\mathrm{F}}(i), \\ c_{\mathrm{F}} \cdot (v^{\mathrm{F}}(i-1) - e^{\mathrm{F}}(i)) \geq \chi^{\mathrm{F}}(i-1) - \chi^{\mathrm{F}}(i) &\geq v^{\mathrm{F}}(i-1) - e^{\mathrm{F}}(i), \end{aligned} \qquad i \in [2, t_{\mathrm{F}}], \tag{14}$$

$$\mathrm{bl}(k,i) \geq \delta_\chi^{\mathrm{F}}(\widetilde{t_{\mathrm{C}}}+i) + \chi^{\mathrm{T}}(i,k) - 1, \qquad k \in [1, k_{\mathrm{C}}], i \in [1, t_{\mathrm{T}}], \tag{15}$$

$$\sum_{k \in [1, k_{\mathrm{C}}], i \in [1, t_{\mathrm{T}}]} \mathrm{bl}(k,i) \leq \sum_{i \in [1, t_{\mathrm{T}}]} \delta_\chi^{\mathrm{F}}(\widetilde{t_{\mathrm{C}}}+i), \tag{16}$$

$$\mathrm{bl}_{\mathrm{LB}}(k) \leq \sum_{i \in [1, t_{\mathrm{T}}]} \mathrm{bl}(k,i) \leq \mathrm{bl}_{\mathrm{UB}}(k), \qquad k \in [1, k_{\mathrm{C}}], \tag{17}$$

$$t_{\mathrm{C}} + \sum_{i \in [1, t_{\mathrm{T}}]} v^{\mathrm{T}}(i) + \sum_{i \in [1, t_{\mathrm{F}}]} v^{\mathrm{F}}(i) = \mathrm{n}_G^{\mathrm{int}}. \tag{18}$$

5.3 Constraints for Including Fringe-trees

Recall that $\mathcal{F}(D_\pi)$ denotes the set of chemical rooted trees ψ r-isomorphic to a chemical rooted tree in $\mathcal{T}(\mathbb{C})$ over all chemical graphs $\mathbb{C} \in D_\pi$, where possibly a chemical rooted tree $\psi \in \mathcal{F}(D_\pi)$ consists of a single chemical element $\mathtt{a} \in \Lambda \setminus \{\mathtt{H}\}$.

To express the condition that the ρ-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}: a lower bound on the number $n(\mathbb{C})$ of non-hydrogen atoms in $\mathbb{C}$, where $n_{\mathrm{LB}}, n^* \geq \mathrm{n}^{\mathrm{int}}_{\mathrm{LB}}$;
- $\mathrm{ch}_{\mathrm{LB}}(i), \mathrm{ch}_{\mathrm{UB}}(i) \in [0, n^*]$, $i \in [1, t_{\mathrm{T}}]$: lower and upper bounds on $\mathrm{ht}(\langle T_i \rangle)$ of the tree T_i rooted at a vertex ${v^{\mathrm{C}}}_i$;
- $\mathrm{ch}_{\mathrm{LB}}(k), \mathrm{ch}_{\mathrm{UB}}(k) \in [0, n^*]$, $k \in [1, k_{\mathrm{C}}] = I_{(\geq 2)} \cup I_{(\geq 1)}$: lower and upper bounds on the maximum height $\mathrm{ht}(\langle T \rangle)$ of the tree $T \in \mathcal{F}(P_k)$ rooted at an internal vertex of a path P_k for an edge $a_k \in E_{(\geq 1)} \cup E_{(\geq 2)}$;
- Prepare a coding of the set $\mathcal{F}(D_\pi)$ and let $[\psi]$ denote the coded integer of an element ψ in $\mathcal{F}(D_\pi)$;
- Sets $\mathcal{F}(v) \subseteq \mathcal{F}(D_\pi), v \in V_{\mathrm{C}}$ and $\mathcal{F}_E \subseteq \mathcal{F}(D_\pi)$ of chemical rooted trees T with $\mathrm{ht}(T) \in [1, \rho]$;
- Define $\mathcal{F}^* := \bigcup_{v \in V_{\mathrm{C}}} \mathcal{F}(v) \cup \mathcal{F}_E$, $\mathcal{F}^{\mathrm{C}}_i := \mathcal{F}({v^{\mathrm{C}}}_i)$, $i \in [1, t_{\mathrm{C}}]$, $\mathcal{F}^{\mathrm{T}}_i := \mathcal{F}_E$, $i \in [1, t_{\mathrm{T}}]$ and $\mathcal{F}^{\mathrm{F}}_i := \mathcal{F}_E$, $i \in [1, t_{\mathrm{F}}]$;
- $\mathrm{fc}_{\mathrm{LB}}(\psi), \mathrm{fc}_{\mathrm{UB}}(\psi) \in [0, n^*], \psi \in \mathcal{F}^*$: lower and upper bound functions on the number of interior-vertices v such that $\mathbb{C}[v]$ is r-isomorphic to ψ in $\mathbb{C}$;
- $\mathcal{F}^{\mathrm{X}}_i[p], p \in [1, \rho], \mathrm{X} \in \{\mathrm{C}, \mathrm{T}, \mathrm{F}\}$: the set of chemical rooted trees $T \in \mathcal{F}^{\mathrm{X}}_i$ with $\mathrm{ht}(\langle T \rangle) = p$;
- $n_{\overline{\mathtt{H}}}([\psi]) \in [0, 3^\rho], \psi \in \mathcal{F}^*$: the number $n(\langle \psi \rangle)$ of non-root hydrogen vertices in a chemical rooted tree ψ;
- $\mathrm{ht}_{\overline{\mathtt{H}}}([\psi]) \in [0, \rho], \psi \in \mathcal{F}^*$: the height $\mathrm{ht}(\langle \psi \rangle)$ of the hydrogen-suppressed chemical rooted tree $\langle \psi \rangle$;
- $\mathrm{deg}^{\overline{\mathtt{H}}}_{\mathrm{r}}([\psi]) \in [0, 3], \psi \in \mathcal{F}^*$: the number $\mathrm{deg}_{\mathrm{r}}(\langle \psi \rangle)$ of non-hydrogen children of the root r of a chemical rooted tree ψ;
- $\mathrm{deg}^{\mathrm{hyd}}_{\mathrm{r}}([\psi]) \in [0, 3], \psi \in \mathcal{F}^*$: the number $\mathrm{deg}_{\mathrm{r}}(\psi) - \mathrm{deg}_{\mathrm{r}}(\langle \psi \rangle)$ of hydrogen children of the root r of a chemical rooted tree ψ;
- $\mathrm{v}_{\mathrm{ion}}(\psi) \in [-3, +3], \psi \in \mathcal{F}^*$: the ion-valence of the root in ψ;
- $\mathrm{ac}^{\mathrm{lf}}_\nu(\psi), \nu \in \Gamma^{\mathrm{lf}}_{\mathrm{ac}}$: the frequency of leaf-edges with adjacency-configuration ν in ψ;
- $\mathrm{ac}^{\mathrm{lf}}_{\mathrm{LB}}, \mathrm{ac}^{\mathrm{lf}}_{\mathrm{UB}} : \Gamma^{\mathrm{lf}}_{\mathrm{ac}} \to [0, n^*]$: lower and upper bound functions on the number of leaf-edges uv in ac_{C} with adjacency-configuration ν;

variables:

- $n_G \in [n_{\mathrm{LB}}, n^*]$: the number $n(\mathbb{C})$ of non-hydrogen atoms in $\mathbb{C}$;
- $v^{\mathrm{X}}(i) \in [0, 1], i \in [1, t_{\mathrm{X}}]$, $\mathrm{X} \in \{\mathrm{T}, \mathrm{F}\}$: $v^{\mathrm{X}}(i) = 1 \Leftrightarrow$ vertex ${v^{\mathrm{X}}}_i$ is used in $\mathbb{C}$;

- $\delta_{\rm fr}^{\rm X}(i,[\psi]) \in [0,1], i \in [1,t_{\rm X}], \psi \in \mathcal{F}_i^{\rm X}, {\rm X} \in \{{\rm C},{\rm T},{\rm F}\}$: $\delta_{\rm fr}^{\rm X}(i,[\psi]) = 1 \Leftrightarrow \psi$ is the ρ-fringe-tree rooted at vertex $v^{\rm X}{}_i$ in $\mathbb{C}$;
- ${\rm fc}([\psi]) \in [{\rm fc}_{\rm LB}(\psi), {\rm fc}_{\rm UB}(\psi)], \psi \in \mathcal{F}^*$: the number of interior-vertices v such that $\mathbb{C}[v]$ is r-isomorphic to ψ in $\mathbb{C}$;
- ${\rm ac}^{\rm lf}([\nu]) \in [{\rm ac}^{\rm lf}_{\rm LB}(\nu), {\rm ac}^{\rm lf}_{\rm UB}(\nu)], \nu \in \Gamma^{\rm lf}_{\rm ac}$: the number of leaf-edge with adjacency-configuration ν in $\mathbb{C}$;
- ${\rm deg}_{\rm X}^{\rm ex}(i) \in [0,3], i \in [1,t_{\rm X}], {\rm X} \in \{{\rm C},{\rm T},{\rm F}\}$: the number of non-hydrogen children of the root of the ρ-fringe-tree rooted at vertex $v^{\rm X}{}_i$ in $\mathbb{C}$;
- ${\rm hyddeg}^{\rm X}(i) \in [0,4]$, $i \in [1,t_{\rm X}]$, ${\rm X} \in \{{\rm C},{\rm T},{\rm F}\}$: the number of hydrogen atoms adjacent to vertex $v^{\rm X}{}_i$ (i.e., ${\rm hyddeg}(v^{\rm X}{}_i)$) in $\mathbb{C} = (H,\alpha,\beta)$;
- ${\rm eledeg}_{\rm X}(i) \in [-3,+3]$, $i \in [1,t_{\rm X}]$, ${\rm X} \in \{{\rm C},{\rm T},{\rm F}\}$: the ion-valence ${\rm v}_{\rm ion}(\psi)$ of vertex $v^{\rm X}{}_i$ (i.e., ${\rm eledeg}_{\rm X}(i) = {\rm v}_{\rm ion}(\psi)$ for the ρ-fringe-tree ψ rooted at $v^{\rm X}{}_i$) in $\mathbb{C} = (H,\alpha,\beta)$;
- $h^{\rm X}(i) \in [0,\rho]$, $i \in [1,t_{\rm X}]$, ${\rm X} \in \{{\rm C},{\rm T},{\rm F}\}$: the height ${\rm ht}(\langle T\rangle)$ of the hydrogen-suppressed chemical rooted tree $\langle T\rangle$ of the ρ-fringe-tree T rooted at vertex $v^{\rm X}{}_i$ in $\mathbb{C}$;
- $\sigma(k,i) \in [0,1]$, $k \in [1,k_{\rm C}] = I_{(\geq 2)} \cup I_{(\geq 1)}, i \in [1,t_{\rm T}]$: $\sigma(k,i) = 1 \Leftrightarrow$ the ρ-fringe-tree T_v rooted at vertex $v = {\rm v}^{\rm T}{}_i$ with color k has the largest height ${\rm ht}(\langle \mathcal{T}_v\rangle)$ among such trees $T_v, v \in V_{\rm T}$;

constraints:

$$\begin{aligned} \sum_{\psi \in \mathcal{F}_i^{\rm C}} \delta_{\rm fr}^{\rm C}(i,[\psi]) &= 1, & i &\in [1,t_{\rm C}], \\ \sum_{\psi \in \mathcal{F}_i^{\rm X}} \delta_{\rm fr}^{\rm X}(i,[\psi]) &= v^{\rm X}(i), & i &\in [1,t_{\rm X}], {\rm X} \in \{{\rm T},{\rm F}\}, \end{aligned} \tag{19}$$

$$\begin{aligned} \sum_{\psi \in \mathcal{F}_i^{\rm X}} {\rm deg}_{\rm r}^{\overline{\rm H}}([\psi]) \cdot \delta_{\rm fr}^{\rm X}(i,[\psi]) &= {\rm deg}_{\rm X}^{\rm ex}(i), \\ \sum_{\psi \in \mathcal{F}_i^{\rm X}} {\rm deg}_{\rm r}^{\rm hyd}([\psi]) \cdot \delta_{\rm fr}^{\rm X}(i,[\psi]) &= {\rm hyddeg}^{\rm X}(i), \\ \sum_{\psi \in \mathcal{F}_i^{\rm X}} {\rm v}_{\rm ion}([\psi]) \cdot \delta_{\rm fr}^{\rm X}(i,[\psi]) &= {\rm eledeg}_{\rm X}(i), & i &\in [1,t_{\rm X}], {\rm X} \in \{{\rm C},{\rm T},{\rm F}\}, \end{aligned} \tag{20}$$

$$\sum_{\psi \in \mathcal{F}_i^{\rm F}[\rho]} \delta_{\rm fr}^{\rm F}(i,[\psi]) \geq v^{\rm F}(i) - e^{\rm F}(i+1), \qquad i \in [1,t_{\rm F}]\ (e^{\rm F}(t_{\rm F}+1) = 0), \tag{21}$$

$$\sum_{\psi \in \mathcal{F}_i^{\rm X}} {\rm ht}_{\overline{\rm H}}([\psi]) \cdot \delta_{\rm fr}^{\rm X}(i,[\psi]) = h^{\rm X}(i), \qquad i \in [1,t_{\rm X}], {\rm X} \in \{{\rm C},{\rm T},{\rm F}\}, \tag{22}$$

$$\sum_{\substack{\psi \in \mathcal{F}_i^{\rm X} \\ i \in [1,t_{\rm X}], {\rm X} \in \{{\rm C},{\rm T},{\rm F}\}}} n_{\overline{\rm H}}([\psi]) \cdot \delta_{\rm fr}^{\rm X}(i,[\psi]) + \sum_{i \in [1,t_{\rm X}], {\rm X} \in \{{\rm T},{\rm F}\}} v^{\rm X}(i) + t_{\rm C} = n_G, \tag{23}$$

$$\sum_{i\in[1,t_{\mathrm{X}}],\mathrm{X}\in\{\mathrm{C,T,F}\}} \delta^{\mathrm{X}}_{\mathrm{fr}}(i,[\psi]) = \mathrm{fc}([\psi]), \qquad \psi\in\mathcal{F}^*, \tag{24}$$

$$\sum_{\psi\in\mathcal{F}^{\mathrm{X}}_i,i\in[1,t_{\mathrm{X}}],\mathrm{X}\in\{\mathrm{C,T,F}\}} \mathrm{ac}^{\mathrm{lf}}_{\nu}(\psi)\cdot\delta^{\mathrm{X}}_{\mathrm{fr}}(i,[\psi]) = \mathrm{ac}^{\mathrm{lf}}([\nu]), \qquad \nu\in\Gamma^{\mathrm{lf}}_{\mathrm{ac}}, \tag{25}$$

$$\begin{aligned} h^{\mathrm{C}}(i) \geq \mathrm{ch}_{\mathrm{LB}}(i) - n^*\cdot\delta^{\mathrm{F}}_{\chi}(i), \quad \mathrm{clr}^{\mathrm{F}}(i)+\rho \geq \mathrm{ch}_{\mathrm{LB}}(i),\\ h^{\mathrm{C}}(i) \leq \mathrm{ch}_{\mathrm{UB}}(i), \quad \mathrm{clr}^{\mathrm{F}}(i)+\rho \leq \mathrm{ch}_{\mathrm{UB}}(i) + n^*\cdot(1-\delta^{\mathrm{F}}_{\chi}(i)), \end{aligned} \qquad i\in[1,\widetilde{t_{\mathrm{C}}}], \tag{26}$$

$$\mathrm{ch}_{\mathrm{LB}}(i) \leq h^{\mathrm{C}}(i) \leq \mathrm{ch}_{\mathrm{UB}}(i), \qquad i\in[\widetilde{t_{\mathrm{C}}}+1,t_{\mathrm{C}}], \tag{27}$$

$$\begin{aligned} h^{\mathrm{T}}(i) \leq \mathrm{ch}_{\mathrm{UB}}(k) + n^*\cdot(\delta^{\mathrm{F}}_{\chi}(\widetilde{t_{\mathrm{C}}}+i)+1-\chi^{\mathrm{T}}(i,k)),\\ \mathrm{clr}^{\mathrm{F}}(\widetilde{t_{\mathrm{C}}}+i)+\rho \leq \mathrm{ch}_{\mathrm{UB}}(k) + n^*\cdot(2-\delta^{\mathrm{F}}_{\chi}(\widetilde{t_{\mathrm{C}}}+i)-\chi^{\mathrm{T}}(i,k)), \end{aligned} \qquad k\in[1,k_{\mathrm{C}}], i\in[1,t_{\mathrm{T}}], \tag{28}$$

$$\sum_{i\in[1,t_{\mathrm{T}}]} \sigma(k,i) = \delta^{\mathrm{T}}_{\chi}(k), \qquad k\in[1,k_{\mathrm{C}}], \tag{29}$$

$$\begin{aligned} \chi^{\mathrm{T}}(i,k) \geq \sigma(k,i),\\ h^{\mathrm{T}}(i) \geq \mathrm{ch}_{\mathrm{LB}}(k) - n^*\cdot(\delta^{\mathrm{F}}_{\chi}(\widetilde{t_{\mathrm{C}}}+i)+1-\sigma(k,i)),\\ \mathrm{clr}^{\mathrm{F}}(\widetilde{t_{\mathrm{C}}}+i)+\rho \geq \mathrm{ch}_{\mathrm{LB}}(k) - n^*\cdot(2-\delta^{\mathrm{F}}_{\chi}(\widetilde{t_{\mathrm{C}}}+i)-\sigma(k,i)), \end{aligned} \qquad k\in[1,k_{\mathrm{C}}], i\in[1,t_{\mathrm{T}}]. \tag{30}$$

5.4 Descriptor for the Number of Specified Degree

We include constraints to compute descriptors for degrees in $\mathbb{C}$.

variables:

- $\deg^{\mathrm{X}}(i)\in[0,4]$, $i\in[1,t_{\mathrm{X}}]$, $\mathrm{X}\in\{\mathrm{C,T,F}\}$: the number of non-hydrogen atoms adjacent to vertex $v=v^{\mathrm{X}}{}_i$ (i.e., $\deg_{\langle\mathbb{C}\rangle}(v)=\deg_H(v)-\mathrm{hyddeg}_{\mathbb{C}}(v)$) in $\mathbb{C}=(H,\alpha,\beta)$;
- $\deg_{\mathrm{CT}}(i)\in[0,4]$, $i\in[1,t_{\mathrm{C}}]$: the number of edges from vertex $v^{\mathrm{C}}{}_i$ to vertices $v^{\mathrm{T}}{}_j$, $j\in[1,t_{\mathrm{T}}]$;
- $\deg_{\mathrm{TC}}(i)\in[0,4]$, $i\in[1,t_{\mathrm{C}}]$: the number of edges from vertices $v^{\mathrm{T}}{}_j$, $j\in[1,t_{\mathrm{T}}]$ to vertex $v^{\mathrm{C}}{}_i$;
- $\delta^{\mathrm{C}}_{\mathrm{dg}}(i,d)\in[0,1]$, $i\in[1,t_{\mathrm{C}}]$, $d\in[1,4]$, $\delta^{\mathrm{X}}_{\mathrm{dg}}(i,d)\in[0,1]$, $i\in[1,t_{\mathrm{X}}]$, $d\in[0,4]$, $\mathrm{X}\in\{\mathrm{T,F}\}$: $\delta^{\mathrm{X}}_{\mathrm{dg}}(i,d)=1 \Leftrightarrow \deg^{\mathrm{X}}(i)+\mathrm{hyddeg}^{\mathrm{X}}(i)=d$;
- $\mathrm{dg}(d)\in[\mathrm{dg}_{\mathrm{LB}}(d),\mathrm{dg}_{\mathrm{UB}}(d)]$, $d\in[1,4]$: the number of interior-vertices v with $\deg_H(v^{\mathrm{X}}{}_i)=d$ in $\mathbb{C}=(H,\alpha,\beta)$;

- $\mathrm{deg}^{\mathrm{int}}_{\mathrm{C}}(i) \in [1,4],\ i \in [1,t_{\mathrm{C}}],\ \mathrm{deg}^{\mathrm{int}}_{\mathrm{X}}(i) \in [0,4],\ i \in [1,t_{\mathrm{X}}], \mathrm{X} \in \{\mathrm{T},\mathrm{F}\}$: the interior-degree $\mathrm{deg}_{H^{\mathrm{int}}}(v^{\mathrm{X}}{}_i)$ in the interior $H^{\mathrm{int}} = (V^{\mathrm{int}}(\mathbb{C}), E^{\mathrm{int}}(\mathbb{C}))$ of $\mathbb{C}$; i.e., the number of interior-edges incident to vertex $v^{\mathrm{X}}{}_i$;
- $\delta^{\mathrm{int}}_{\mathrm{dg,C}}(i,d) \in [0,1],\ i \in [1,t_{\mathrm{C}}],\ d \in [1,4],\ \delta^{\mathrm{int}}_{\mathrm{dg,X}}(i,d) \in [0,1],\ i \in [1,t_{\mathrm{X}}],\ d \in [0,4],\ \mathrm{X} \in \{\mathrm{T},\mathrm{F}\}$: $\delta^{\mathrm{int}}_{\mathrm{dg,X}}(i,d) = 1 \Leftrightarrow \mathrm{deg}^{\mathrm{int}}_{\mathrm{X}}(i) = d$;
- $\mathrm{dg}^{\mathrm{int}}(d) \in [\mathrm{dg}_{\mathrm{LB}}(d), \mathrm{dg}_{\mathrm{UB}}(d)],\ d \in [1,4]$: the number of interior-vertices v with the interior-degree $\mathrm{deg}_{H^{\mathrm{int}}}(v) = d$ in the interior $H^{\mathrm{int}} = (V^{\mathrm{int}}(\mathbb{C}), E^{\mathrm{int}}(\mathbb{C}))$ of $\mathbb{C} = (H, \alpha, \beta)$.

constraints:

$$\sum_{k \in I^{+}_{(\geq 2)}(i) \cup I^{+}_{(\geq 1)}(i)} \delta^{\mathrm{T}}_{\chi}(k) = \mathrm{deg}_{\mathrm{CT}}(i), \quad \sum_{k \in I^{-}_{(\geq 2)}(i) \cup I^{-}_{(\geq 1)}(i)} \delta^{\mathrm{T}}_{\chi}(k) = \mathrm{deg}_{\mathrm{TC}}(i), \qquad i \in [1,t_{\mathrm{C}}], \tag{31}$$

$$\widetilde{\mathrm{deg}}^{-}_{\mathrm{C}}(i) + \widetilde{\mathrm{deg}}^{+}_{\mathrm{C}}(i) + \mathrm{deg}_{\mathrm{CT}}(i) + \mathrm{deg}_{\mathrm{TC}}(i) + \delta^{\mathrm{F}}_{\chi}(i) = \mathrm{deg}^{\mathrm{int}}_{\mathrm{C}}(i), \qquad i \in [1, \widetilde{t_{\mathrm{C}}}], \tag{32}$$

$$\widetilde{\mathrm{deg}}^{-}_{\mathrm{C}}(i) + \widetilde{\mathrm{deg}}^{+}_{\mathrm{C}}(i) + \mathrm{deg}_{\mathrm{CT}}(i) + \mathrm{deg}_{\mathrm{TC}}(i) = \mathrm{deg}^{\mathrm{int}}_{\mathrm{C}}(i), \qquad i \in [\widetilde{t_{\mathrm{C}}}+1, t_{\mathrm{C}}], \tag{33}$$

$$\mathrm{deg}^{\mathrm{int}}_{\mathrm{C}}(i) + \mathrm{deg}^{\mathrm{ex}}_{\mathrm{C}}(i) = \mathrm{deg}^{\mathrm{C}}(i), \qquad i \in [1,t_{\mathrm{C}}], \tag{34}$$

$$\sum_{\psi \in \mathcal{F}^{\mathrm{C}}_{i}[\rho]} \delta^{\mathrm{C}}_{\mathrm{fr}}(i,[\psi]) \geq 2 - \mathrm{deg}^{\mathrm{int}}_{\mathrm{C}}(i) \qquad i \in [1,t_{\mathrm{C}}], \tag{35}$$

$$\begin{aligned} 2v^{\mathrm{T}}(i) + \delta^{\mathrm{F}}_{\chi}(\widetilde{t_{\mathrm{C}}} + i) &= \mathrm{deg}^{\mathrm{int}}_{\mathrm{T}}(i), \\ \mathrm{deg}^{\mathrm{int}}_{\mathrm{T}}(i) + \mathrm{deg}^{\mathrm{ex}}_{\mathrm{T}}(i) &= \mathrm{deg}^{\mathrm{T}}(i), \qquad i \in [1,t_{\mathrm{T}}]\ (e^{\mathrm{T}}(1) = e^{\mathrm{T}}(t_{\mathrm{T}}+1) = 0), \end{aligned} \tag{36}$$

$$\begin{aligned} v^{\mathrm{F}}(i) + e^{\mathrm{F}}(i+1) &= \mathrm{deg}^{\mathrm{int}}_{\mathrm{F}}(i), \\ \mathrm{deg}^{\mathrm{int}}_{\mathrm{F}}(i) + \mathrm{deg}^{\mathrm{ex}}_{\mathrm{F}}(i) &= \mathrm{deg}^{\mathrm{F}}(i), \qquad i \in [1,t_{\mathrm{F}}]\ (e^{\mathrm{F}}(1) = e^{\mathrm{F}}(t_{\mathrm{F}}+1) = 0), \end{aligned} \tag{37}$$

$$\begin{aligned} &\sum_{d \in [0,4]} \delta^{\mathrm{X}}_{\mathrm{dg}}(i,d) = 1, \quad \sum_{d \in [1,4]} d \cdot \delta^{\mathrm{X}}_{\mathrm{dg}}(i,d) = \mathrm{deg}^{\mathrm{X}}(i) + \mathrm{hyddeg}^{\mathrm{X}}(i), \\ &\sum_{d \in [0,4]} \delta^{\mathrm{int}}_{\mathrm{dg,X}}(i,d) = 1, \quad \sum_{d \in [1,4]} d \cdot \delta^{\mathrm{int}}_{\mathrm{dg,X}}(i,d) = \mathrm{deg}^{\mathrm{int}}_{\mathrm{X}}(i), \quad i \in [1,t_{\mathrm{X}}], \mathrm{X} \in \{\mathrm{T},\mathrm{C},\mathrm{F}\}, \end{aligned} \tag{38}$$

$$\begin{aligned} &\sum_{i \in [1,t_{\mathrm{C}}]} \delta^{\mathrm{C}}_{\mathrm{dg}}(i,d) + \sum_{i \in [1,t_{\mathrm{T}}]} \delta^{\mathrm{T}}_{\mathrm{dg}}(i,d) + \sum_{i \in [1,t_{\mathrm{F}}]} \delta^{\mathrm{F}}_{\mathrm{dg}}(i,d) = \mathrm{dg}(d), \\ &\sum_{i \in [1,t_{\mathrm{C}}]} \delta^{\mathrm{int}}_{\mathrm{dg,C}}(i,d) + \sum_{i \in [1,t_{\mathrm{T}}]} \delta^{\mathrm{int}}_{\mathrm{dg,T}}(i,d) + \sum_{i \in [1,t_{\mathrm{F}}]} \delta^{\mathrm{int}}_{\mathrm{dg,F}}(i,d) = \mathrm{dg}^{\mathrm{int}}(d), \qquad d \in [1,4]. \end{aligned} \tag{39}$$

5.5 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 H and include necessary constraints for the variables to satisfy in H.

constants:

- $\beta_{\mathrm{r}}([\psi])$: the sum $\beta_\psi(r)$ of bond-multiplicities of edges incident to the root r of a chemical rooted tree $\psi \in \mathcal{F}^*$;

variables:

- $\beta^{\mathrm{X}}(i) \in [0,3]$, $i \in [2, t_{\mathrm{X}}]$, $\mathrm{X} \in \{\mathrm{T}, \mathrm{F}\}$: the bond-multiplicity of edge ${e^{\mathrm{X}}}_i$ in $\mathbb{C}$;
- $\beta^{\mathrm{C}}(i) \in [0,3]$, $i \in [\widetilde{k_{\mathrm{C}}}+1, m_{\mathrm{C}}] = I_{(\geq 1)} \cup I_{(0/1)} \cup I_{(=1)}$: the bond-multiplicity of edge $a_i \in E_{(\geq 1)} \cup E_{(0/1)} \cup E_{(=1)}$ in $\mathbb{C}$;
- $\beta^{\mathrm{CT}}(k), \beta^{\mathrm{TC}}(k) \in [0,3]$, $k \in [1, k_{\mathrm{C}}] = I_{(\geq 2)} \cup I_{(\geq 1)}$: the bond-multiplicity of the first (resp., last) edge of the pure path P_k in $\mathbb{C}$;
- $\beta^{*\mathrm{F}}(c) \in [0,3], c \in [1, c_{\mathrm{F}} = \widetilde{t_{\mathrm{C}}} + t_{\mathrm{T}}]$: the bond-multiplicity of the first edge of the leaf path Q_c rooted at vertex ${v^{\mathrm{C}}}_c, c \leq \widetilde{t_{\mathrm{C}}}$ or ${v^{\mathrm{T}}}_{c-\widetilde{t_{\mathrm{C}}}}, c > \widetilde{t_{\mathrm{C}}}$ in $\mathbb{C}$;
- $\beta^{\mathrm{X}}_{\mathrm{ex}}(i) \in [0,4], i \in [1, t_{\mathrm{X}}], \mathrm{X} \in \{\mathrm{C}, \mathrm{T}, \mathrm{F}\}$: the sum $\beta_{\mathbb{C}[v]}(v)$ of bond-multiplicities of edges in the ρ-fringe-tree $\mathbb{C}[v]$ rooted at interior-vertex $v = {v^{\mathrm{X}}}_i$;
- $\delta^{\mathrm{X}}_\beta(i,m) \in [0,1]$, $i \in [2, t_{\mathrm{X}}]$, $m \in [0,3]$, $\mathrm{X} \in \{\mathrm{T}, \mathrm{F}\}$: $\delta^{\mathrm{X}}_\beta(i,m) = 1 \Leftrightarrow \beta^{\mathrm{X}}(i) = m$;
- $\delta^{\mathrm{C}}_\beta(i,m) \in [0,1]$, $i \in [\widetilde{k_{\mathrm{C}}}, m_{\mathrm{C}}] = I_{(\geq 1)} \cup I_{(0/1)} \cup I_{(=1)}$, $m \in [0,3]$: $\delta^{\mathrm{C}}_\beta(i,m) = 1 \Leftrightarrow \beta^{\mathrm{C}}(i) = m$;
- $\delta^{\mathrm{CT}}_\beta(k,m), \delta^{\mathrm{TC}}_\beta(k,m) \in [0,1]$, $k \in [1, k_{\mathrm{C}}] = I_{(\geq 2)} \cup I_{(\geq 1)}$, $m \in [0,3]$: $\delta^{\mathrm{CT}}_\beta(k,m) = 1$ (resp., $\delta^{\mathrm{TC}}_\beta(k,m) = 1$) $\Leftrightarrow \beta^{\mathrm{CT}}(k) = m$ (resp., $\beta^{\mathrm{TC}}(k) = m$);
- $\delta^{*\mathrm{F}}_\beta(c,m) \in [0,1]$, $c \in [1, c_{\mathrm{F}}]$, $m \in [0,3], \mathrm{X} \in \{\mathrm{C}, \mathrm{T}\}$: $\delta^{*\mathrm{F}}_\beta(c,m) = 1 \Leftrightarrow \beta^{*\mathrm{F}}(c) = m$;
- $\mathrm{bd}^{\mathrm{int}}(m) \in [0, 2\mathrm{n}^{\mathrm{int}}_{\mathrm{UB}}]$, $m \in [1,3]$: the number of interior-edges with bond-multiplicity m in $\mathbb{C}$;
- $\mathrm{bd}_{\mathrm{X}}(m) \in [0, 2\mathrm{n}^{\mathrm{int}}_{\mathrm{UB}}], \mathrm{X} \in \{\mathrm{C}, \mathrm{T}, \mathrm{CT}, \mathrm{TC}\}$, $\mathrm{bd}_{\mathrm{X}}(m) \in [0, 2\mathrm{n}^{\mathrm{int}}_{\mathrm{UB}}], \mathrm{X} \in \{\mathrm{F}, \mathrm{CF}, \mathrm{TF}\}$, $m \in [1,3]$: the number of interior-edges $e \in E_{\mathrm{X}}$ with bond-multiplicity m in $\mathbb{C}$;

constraints:

$$e^{\mathrm{C}}(i) \leq \beta^{\mathrm{C}}(i) \leq 3e^{\mathrm{C}}(i), i \in [\widetilde{k_{\mathrm{C}}}+1, m_{\mathrm{C}}] = I_{(\geq 1)} \cup I_{(0/1)} \cup I_{(=1)}, \tag{40}$$

$$e^{\mathrm{X}}(i) \leq \beta^{\mathrm{X}}(i) \leq 3e^{\mathrm{X}}(i), \qquad i \in [2, t_{\mathrm{X}}], \mathrm{X} \in \{\mathrm{T}, \mathrm{F}\}, \tag{41}$$

$$\delta^{\mathrm{T}}_\chi(k) \leq \beta^{\mathrm{CT}}(k) \leq 3\delta^{\mathrm{T}}_\chi(k), \quad \delta^{\mathrm{T}}_\chi(k) \leq \beta^{\mathrm{TC}}(k) \leq 3\delta^{\mathrm{T}}_\chi(k), \qquad k \in [1, k_{\mathrm{C}}], \tag{42}$$

$$\delta^{\mathrm{F}}_\chi(c) \leq \beta^{\mathrm{XF}}(c) \leq 3\delta^{\mathrm{F}}_\chi(c), \qquad c \in [1, c_{\mathrm{F}}] \tag{43}$$

$$\sum_{m\in[0,3]} \delta_\beta^{\mathrm{X}}(i,m) = 1, \quad \sum_{m\in[0,3]} m\cdot\delta_\beta^{\mathrm{X}}(i,m) = \beta^{\mathrm{X}}(i), \qquad i\in[2,t_{\mathrm{X}}], \mathrm{X}\in\{\mathrm{T},\mathrm{F}\}, \tag{44}$$

$$\sum_{m\in[0,3]} \delta_\beta^{\mathrm{C}}(i,m) = 1, \quad \sum_{m\in[0,3]} m\cdot\delta_\beta^{\mathrm{C}}(i,m) = \beta^{\mathrm{C}}(i), \qquad i\in[\widetilde{k_{\mathrm{C}}}+1, m_{\mathrm{C}}], \tag{45}$$

$$\begin{aligned}
&\sum_{m\in[0,3]} \delta_\beta^{\mathrm{CT}}(k,m) = 1, \quad \sum_{m\in[0,3]} m\cdot\delta_\beta^{\mathrm{CT}}(k,m) = \beta^{\mathrm{CT}}(k), && k\in[1,k_{\mathrm{C}}],\\
&\sum_{m\in[0,3]} \delta_\beta^{\mathrm{TC}}(k,m) = 1, \quad \sum_{m\in[0,3]} m\cdot\delta_\beta^{\mathrm{TC}}(k,m) = \beta^{\mathrm{TC}}(k), && k\in[1,k_{\mathrm{C}}],\\
&\sum_{m\in[0,3]} \delta_\beta^{*\mathrm{F}}(c,m) = 1, \quad \sum_{m\in[0,3]} m\cdot\delta_\beta^{*\mathrm{F}}(c,m) = \beta^{*\mathrm{F}}(c), && c\in[1,c_{\mathrm{F}}],
\end{aligned} \tag{46}$$

$$\sum_{\psi\in\mathcal{F}_i^{\mathrm{X}}} \beta_{\mathrm{r}}([\psi])\cdot\delta_{\mathrm{fr}}^{\mathrm{X}}(i,[\psi]) = \beta_{\mathrm{ex}}^{\mathrm{X}}(i), \qquad i\in[1,t_{\mathrm{X}}], \mathrm{X}\in\{\mathrm{C},\mathrm{T},\mathrm{F}\}, \tag{47}$$

$$\begin{aligned}
\sum_{i\in[\widetilde{k_{\mathrm{C}}}+1,m_{\mathrm{C}}]} \delta_\beta^{\mathrm{C}}(i,m) = \mathrm{bd}_{\mathrm{C}}(m), \quad & \sum_{i\in[2,t_{\mathrm{T}}]} \delta_\beta^{\mathrm{T}}(i,m) = \mathrm{bd}_{\mathrm{T}}(m),\\
\sum_{k\in[1,k_{\mathrm{C}}]} \delta_\beta^{\mathrm{CT}}(k,m) = \mathrm{bd}_{\mathrm{CT}}(m), \quad & \sum_{k\in[1,k_{\mathrm{C}}]} \delta_\beta^{\mathrm{TC}}(k,m) = \mathrm{bd}_{\mathrm{TC}}(m),\\
\sum_{i\in[2,t_{\mathrm{F}}]} \delta_\beta^{\mathrm{F}}(i,m) = \mathrm{bd}_{\mathrm{F}}(m), \quad & \sum_{c\in[1,\widetilde{t_{\mathrm{C}}}]} \delta_\beta^{*\mathrm{F}}(c,m) = \mathrm{bd}_{\mathrm{CF}}(m),\\
& \sum_{c\in[\widetilde{t_{\mathrm{C}}}+1,c_{\mathrm{F}}]} \delta_\beta^{*\mathrm{F}}(c,m) = \mathrm{bd}_{\mathrm{TF}}(m),\\
\mathrm{bd}_{\mathrm{C}}(m)+\mathrm{bd}_{\mathrm{T}}(m)+\mathrm{bd}_{\mathrm{F}}(m)+\mathrm{bd}_{\mathrm{CT}}(m)+\mathrm{bd}_{\mathrm{TC}}(m)&+\mathrm{bd}_{\mathrm{TF}}(m)+\mathrm{bd}_{\mathrm{CF}}(m) = \mathrm{bd}^{\mathrm{int}}(m),\\
& m\in[1,3].
\end{aligned} \tag{48}$$

5.6 Assigning Chemical Elements and Valence Condition

We include constraints so that each vertex v in a selected graph H satisfies the valence condition; i.e., $\beta_{\mathbb{C}}(v) = \mathrm{val}(\alpha(v)) + \mathrm{eledeg}_{\mathbb{C}}(v)$, where $\mathrm{eledeg}_{\mathbb{C}}(v) = \mathrm{v_{ion}}(\psi)$ for the ρ-fringe-tree $\mathbb{C}[v]$ r-isomorphic to ψ. With these constraints, a chemical graph $\mathbb{C} = (H,\alpha,\beta)$ on a selected subgraph H will be constructed.

constants:

- Subsets $\Lambda^{\mathrm{int}} \subseteq \Lambda\setminus\{\mathtt{H}\}, \Lambda^{\mathrm{ex}} \subseteq \Lambda$ of chemical elements, where we denote by $[\mathtt{e}]$ (resp., $[\mathtt{e}]^{\mathrm{int}}$ and $[\mathtt{e}]^{\mathrm{ex}}$) of a standard encoding of an element $\mathtt{e}$ in the set Λ (resp., $\Lambda_\epsilon^{\mathrm{int}}$ and $\Lambda_\epsilon^{\mathrm{ex}}$);
- A valence function: $\mathrm{val}: \Lambda \to [1,6]$;
- A function $\mathrm{mass}^*: \Lambda\to\mathbb{Z}$ (we let $\mathrm{mass}(\mathtt{a})$ denote the observed mass of a chemical element $\mathtt{a}\in\Lambda$, and define $\mathrm{mass}^*(\mathtt{a}) \triangleq \lfloor 10\cdot\mathrm{mass}(\mathtt{a})\rfloor$);

- Subsets $\Lambda^*(i) \subseteq \Lambda^{\rm int}$, $i \in [1, t_{\rm C}]$;
- $\mathrm{na}_{\rm LB}(\mathtt{a}), \mathrm{na}_{\rm UB}(\mathtt{a}) \in [0, n^*]$, $\mathtt{a} \in \Lambda$: lower and upper bounds on the number of vertices v with $\alpha(v) = \mathtt{a}$;
- $\mathrm{na}^{\rm int}_{\rm LB}(\mathtt{a}), \mathrm{na}^{\rm int}_{\rm UB}(\mathtt{a}) \in [0, n^*]$, $\mathtt{a} \in \Lambda^{\rm int}$: lower and upper bounds on the number of interior-vertices v with $\alpha(v) = \mathtt{a}$;
- $\alpha_{\rm r}([\psi]) \in [\Lambda^{\rm ex}], \in \mathcal{F}^*$: the chemical element $\alpha(r)$ of the root r of ψ;
- $\mathrm{na}^{\rm ex}_{\mathtt{a}}([\psi]) \in [0, n^*]$, $\mathtt{a} \in \Lambda^{\rm ex}, \psi \in \mathcal{F}^*$: the frequency of chemical element $\mathtt{a}$ in the set of non-rooted vertices in ψ, where possibly $\mathtt{a} = \mathtt{H}$;
- A positive integer $\mathrm{M} \in \mathbb{Z}_+$: an upper bound for the average $\overline{\mathrm{ms}}(\mathbb{C})$ of mass* over all atoms in $\mathbb{C}$;

variables:

- $\beta^{\rm CT}(i), \beta^{\rm TC}(i) \in [0, 3], i \in [1, t_{\rm T}]$: the bond-multiplicity of edge $e^{\rm CT}{}_{j,i}$ (resp., $e^{\rm TC}{}_{j,i}$) if one exists;
- $\beta^{\rm CF}(i), \beta^{\rm TF}(i) \in [0, 3], i \in [1, t_{\rm F}]$: the bond-multiplicity of $e^{\rm CF}{}_{j,i}$ (resp., $e^{\rm TF}{}_{j,i}$) if one exists;
- $\alpha^{\rm X}(i) \in [\Lambda^{\rm int}_{\epsilon}], \delta^{\rm X}_{\alpha}(i, [\mathtt{a}]^{\rm int}) \in [0, 1], \mathtt{a} \in \Lambda^{\rm int}_{\epsilon}, i \in [1, t_{\rm X}], \mathrm{X} \in \{\mathrm{C}, \mathrm{T}, \mathrm{F}\}$: $\alpha^{\rm X}(i) = [\mathtt{a}]^{\rm int} \geq 1$ (resp., $\alpha^{\rm X}(i) = 0$) $\Leftrightarrow \delta^{\rm X}_{\alpha}(i, [\mathtt{a}]^{\rm int}) = 1$ (resp., $\delta^{\rm X}_{\alpha}(i, 0) = 0$) $\Leftrightarrow \alpha(v^{\rm X}{}_i) = \mathtt{a} \in \Lambda$ (resp., vertex $v^{\rm X}{}_i$ is not used in $\mathbb{C}$);
- $\delta^{\rm X}_{\alpha}(i, [\mathtt{a}]^{\rm int}) \in [0, 1], i \in [1, t_{\rm X}], \mathtt{a} \in \Lambda^{\rm int}, \mathrm{X} \in \{\mathrm{C}, \mathrm{T}, \mathrm{F}\}$: $\delta^{\rm X}_{\alpha}(i, [\mathtt{a}]^{\rm t}) = 1 \Leftrightarrow \alpha(v^{\rm X}{}_i) = \mathtt{a}$;
- $\mathrm{Mass} \in \mathbb{Z}_+$: $\sum_{v \in V(H)} \mathrm{mass}^*(\alpha(v))$;
- $\overline{\mathrm{ms}} \in \mathbb{R}_+$: $\sum_{v \in V(H)} \mathrm{mass}^*(\alpha(v))/|V(H)|$;
- $\delta_{\rm atm}(i) \in [0, 1], i \in [n_{\rm LB} + \mathrm{na}_{\rm LB}(\mathtt{H}), n^* + \mathrm{na}_{\rm UB}(\mathtt{H})]$: $\delta_{\rm atm}(i) = 1 \Leftrightarrow |V(H)| = i$;
- $\mathrm{na}([\mathtt{a}]) \in [\mathrm{na}_{\rm LB}(\mathtt{a}), \mathrm{na}_{\rm UB}(\mathtt{a})]$, $\mathtt{a} \in \Lambda$: the number of vertices $v \in V(H)$ with $\alpha(v) = \mathtt{a}$, where possibly $\mathtt{a} = \mathtt{H}$;
- $\mathrm{na}^{\rm int}([\mathtt{a}]^{\rm int}) \in [\mathrm{na}^{\rm int}_{\rm LB}(\mathtt{a}), \mathrm{na}^{\rm int}_{\rm UB}(\mathtt{a})]$, $\mathtt{a} \in \Lambda, \mathrm{X} \in \{\mathrm{C}, \mathrm{T}, \mathrm{F}\}$: the number of interior-vertices $v \in V(\mathbb{C})$ with $\alpha(v) = \mathtt{a}$;
- $\mathrm{na}^{\rm ex}_{\rm X}([\mathtt{a}]^{\rm ex}), \mathrm{na}^{\rm ex}([\mathtt{a}]^{\rm ex}) \in [0, \mathrm{na}_{\rm UB}(\mathtt{a})]$, $\mathtt{a} \in \Lambda$, $\mathrm{X} \in \{\mathrm{C}, \mathrm{T}, \mathrm{F}\}$: the number of exterior-vertices rooted at vertices $v \in V_{\rm X}$ and the number of exterior-vertices v such that $\alpha(v) = \mathtt{a}$;

constraints:

$$\begin{aligned}
\beta^{\rm CT}(k) - 3(e^{\rm T}(i) - \chi^{\rm T}(i,k) + 1) \leq \beta^{\rm CT}(i) \leq \beta^{\rm CT}(k) + 3(e^{\rm T}(i) - \chi^{\rm T}(i,k) + 1), i \in [1, t_{\rm T}], \\
\beta^{\rm TC}(k) - 3(e^{\rm T}(i+1) - \chi^{\rm T}(i,k) + 1) \leq \beta^{\rm TC}(i) \leq \beta^{\rm TC}(k) + 3(e^{\rm T}(i+1) - \chi^{\rm T}(i,k) + 1), i \in [1, t_{\rm T}], \\
k \in [1, k_{\rm C}],
\end{aligned} \tag{49}$$

$$\begin{aligned}
\beta^{*\rm F}(c) - 3(e^{\rm F}(i) - \chi^{\rm F}(i,c) + 1) \leq \beta^{\rm CF}(i) \leq \beta^{*\rm F}(c) + 3(e^{\rm F}(i) - \chi^{\rm F}(i,c) + 1), i \in [1, t_{\rm F}], \quad c \in [1, \widetilde{t_{\rm C}}], \\
\beta^{*\rm F}(c) - 3(e^{\rm F}(i) - \chi^{\rm F}(i,c) + 1) \leq \beta^{\rm TF}(i) \leq \beta^{*\rm F}(c) + 3(e^{\rm F}(i) - \chi^{\rm F}(i,c) + 1), i \in [1, t_{\rm F}], \quad c \in [\widetilde{t_{\rm C}} + 1, c_{\rm F}],
\end{aligned} \tag{50}$$

$$\sum_{\mathtt{a}\in\Lambda^{\mathrm{int}}} \delta_{\alpha}^{\mathrm{C}}(i,[\mathtt{a}]^{\mathrm{int}}) = 1, \quad \sum_{\mathtt{a}\in\Lambda^{\mathrm{int}}} [\mathtt{a}]^{\mathrm{int}} \cdot \delta_{\alpha}^{\mathrm{X}}(i,[\mathtt{a}]^{\mathrm{int}}) = \alpha^{\mathrm{C}}(i), \qquad i\in[1,t_{\mathrm{C}}],$$

$$\sum_{\mathtt{a}\in\Lambda^{\mathrm{int}}} \delta_{\alpha}^{\mathrm{X}}(i,[\mathtt{a}]^{\mathrm{int}}) = v^{\mathrm{X}}(i), \quad \sum_{\mathtt{a}\in\Lambda^{\mathrm{int}}} [\mathtt{a}]^{\mathrm{int}} \cdot \delta_{\alpha}^{\mathrm{X}}(i,[\mathtt{a}]^{\mathrm{int}}) = \alpha^{\mathrm{X}}(i), \qquad i\in[1,t_{\mathrm{X}}], \mathrm{X}\in\{\mathrm{T},\mathrm{F}\}, \tag{51}$$

$$\sum_{\psi\in\mathcal{F}_i^{\mathrm{X}}} \alpha_{\mathrm{r}}([\psi])\cdot\delta_{\mathrm{fr}}^{\mathrm{X}}(i,[\psi]) = \alpha^{\mathrm{X}}(i), \qquad i\in[1,t_{\mathrm{X}}], \mathrm{X}\in\{\mathrm{C},\mathrm{T},\mathrm{F}\}, \tag{52}$$

$$\begin{aligned}\sum_{j\in I_{\mathrm{C}}(i)} \beta^{\mathrm{C}}(j) + \sum_{k\in I^{+}_{(\geq 2)}(i)\cup I^{+}_{(\geq 1)}(i)} \beta^{\mathrm{CT}}(k) + \sum_{k\in I^{-}_{(\geq 2)}(i)\cup I^{-}_{(\geq 1)}(i)} \beta^{\mathrm{TC}}(k) \\ +\beta^{*\mathrm{F}}(i) + \beta^{\mathrm{C}}_{\mathrm{ex}}(i) - \mathrm{eledeg}_{\mathrm{C}}(i) = \sum_{\mathtt{a}\in\Lambda^{\mathrm{int}}} \mathrm{val}(\mathtt{a})\delta_{\alpha}^{\mathrm{C}}(i,[\mathtt{a}]^{\mathrm{int}}), \qquad i\in[1,\widetilde{t_{\mathrm{C}}}],\end{aligned} \tag{53}$$

$$\begin{aligned}\sum_{j\in I_{\mathrm{C}}(i)} \beta^{\mathrm{C}}(j) + \sum_{k\in I^{+}_{(\geq 2)}(i)\cup I^{+}_{(\geq 1)}(i)} \beta^{\mathrm{CT}}(k) + \sum_{k\in I^{-}_{(\geq 2)}(i)\cup I^{-}_{(\geq 1)}(i)} \beta^{\mathrm{TC}}(k) \\ +\beta^{\mathrm{C}}_{\mathrm{ex}}(i) - \mathrm{eledeg}_{\mathrm{C}}(i) = \sum_{\mathtt{a}\in\Lambda^{\mathrm{int}}} \mathrm{val}(\mathtt{a})\delta_{\alpha}^{\mathrm{C}}(i,[\mathtt{a}]^{\mathrm{int}}), \qquad i\in[\widetilde{t_{\mathrm{C}}}+1,t_{\mathrm{C}}],\end{aligned} \tag{54}$$

$$\begin{aligned}\beta^{\mathrm{T}}(i) + \beta^{\mathrm{T}}(i{+}1) + \beta^{\mathrm{T}}_{\mathrm{ex}}(i) + \beta^{\mathrm{CT}}(i) + \beta^{\mathrm{TC}}(i) \\ +\beta^{*\mathrm{F}}(\widetilde{t_{\mathrm{C}}}+i) - \mathrm{eledeg}_{\mathrm{T}}(i) = \sum_{\mathtt{a}\in\Lambda^{\mathrm{int}}} \mathrm{val}(\mathtt{a})\delta_{\alpha}^{\mathrm{T}}(i,[\mathtt{a}]^{\mathrm{int}}), \\ i\in[1,t_{\mathrm{T}}]\ (\beta^{\mathrm{T}}(1)=\beta^{\mathrm{T}}(t_{\mathrm{T}}+1)=0),\end{aligned} \tag{55}$$

$$\begin{aligned}\beta^{\mathrm{F}}(i) + \beta^{\mathrm{F}}(i{+}1) + \beta^{\mathrm{CF}}(i) + \beta^{\mathrm{TF}}(i) \\ +\beta^{\mathrm{F}}_{\mathrm{ex}}(i) - \mathrm{eledeg}_{\mathrm{F}}(i) = \sum_{\mathtt{a}\in\Lambda^{\mathrm{int}}} \mathrm{val}(\mathtt{a})\delta_{\alpha}^{\mathrm{F}}(i,[\mathtt{a}]^{\mathrm{int}}), \\ i\in[1,t_{\mathrm{F}}]\ (\beta^{\mathrm{F}}(1)=\beta^{\mathrm{F}}(t_{\mathrm{F}}+1)=0),\end{aligned} \tag{56}$$

$$\sum_{i\in[1,t_{\mathrm{X}}],i\in[1,t_{\mathrm{X}}]} \delta_{\alpha}^{\mathrm{X}}(i,[\mathtt{a}]^{\mathrm{int}}) = \mathrm{na}_{\mathrm{X}}([\mathtt{a}]^{\mathrm{int}}), \qquad \mathtt{a}\in\Lambda^{\mathrm{int}}, \mathrm{X}\in\{\mathrm{C},\mathrm{T},\mathrm{F}\}, \tag{57}$$

$$\sum_{\psi\in\mathcal{F}_i^{\mathrm{X}}} \mathrm{na}^{\mathrm{ex}}_{\mathtt{a}}([\psi])\cdot\delta_{\mathrm{fr}}^{\mathrm{X}}(i,[\psi]) = \mathrm{na}^{\mathrm{ex}}_{\mathrm{X}}([\mathtt{a}]^{\mathrm{ex}}), \qquad \mathtt{a}\in\Lambda^{\mathrm{ex}}, \mathrm{X}\in\{\mathrm{C},\mathrm{T},\mathrm{F}\}, \tag{58}$$

$$\begin{aligned}
\mathrm{na}_{\mathrm{C}}([\mathtt{a}]^{\mathrm{int}}) + \mathrm{na}_{\mathrm{T}}([\mathtt{a}]^{\mathrm{int}}) + \mathrm{na}_{\mathrm{F}}([\mathtt{a}]^{\mathrm{int}}) = \mathrm{na}^{\mathrm{int}}([\mathtt{a}]^{\mathrm{int}}), &\qquad \mathtt{a} \in \Lambda^{\mathrm{int}}, \\
\sum_{\mathrm{X}\in\{\mathrm{C,T,F}\}} \mathrm{na}^{\mathrm{ex}}_{\mathrm{X}}([\mathtt{a}]^{\mathrm{ex}}) = \mathrm{na}^{\mathrm{ex}}([\mathtt{a}]^{\mathrm{ex}}), &\qquad \mathtt{a} \in \Lambda^{\mathrm{ex}}, \\
\mathrm{na}^{\mathrm{int}}([\mathtt{a}]^{\mathrm{int}}) + \mathrm{na}^{\mathrm{ex}}([\mathtt{a}]^{\mathrm{ex}}) = \mathrm{na}([\mathtt{a}]), &\qquad \mathtt{a} \in \Lambda^{\mathrm{int}} \cap \Lambda^{\mathrm{ex}}, \\
\mathrm{na}^{\mathrm{int}}([\mathtt{a}]^{\mathrm{int}}) = \mathrm{na}([\mathtt{a}]), &\qquad \mathtt{a} \in \Lambda^{\mathrm{int}} \setminus \Lambda^{\mathrm{ex}}, \\
\mathrm{na}^{\mathrm{ex}}([\mathtt{a}]^{\mathrm{ex}}) = \mathrm{na}([\mathtt{a}]), &\qquad \mathtt{a} \in \Lambda^{\mathrm{ex}} \setminus \Lambda^{\mathrm{int}}, \qquad (59)
\end{aligned}$$

$$\sum_{\mathtt{a}\in\Lambda^*(i)} \delta^{\mathrm{C}}_{\alpha}(i, [\mathtt{a}]^{\mathrm{int}}) = 1, \qquad i \in [1, t_{\mathrm{C}}], \tag{60}$$

$$\sum_{\mathtt{a}\in\Lambda} \mathrm{mass}^*(\mathtt{a}) \cdot \mathrm{na}([\mathtt{a}]) = \mathrm{Mass}, \tag{61}$$

$$\sum_{i\in[n_{\mathrm{LB}}+\mathrm{na}_{\mathrm{LB}}(\mathtt{H}), n^*+\mathrm{na}_{\mathrm{UB}}(\mathtt{H})]} \delta_{\mathrm{atm}}(i) = 1, \tag{62}$$

$$\sum_{i\in[n_{\mathrm{LB}}+\mathrm{na}_{\mathrm{LB}}(\mathtt{H}), n^*+\mathrm{na}_{\mathrm{UB}}(\mathtt{H})]} i \cdot \delta_{\mathrm{atm}}(i) = n_G + \mathrm{na}^{\mathrm{ex}}([\mathtt{H}]^{\mathrm{ex}}), \tag{63}$$

$$\mathrm{Mass} - \mathrm{M} \cdot (1 - \delta_{\mathrm{atm}}(i)) \le i \cdot \overline{\mathrm{ms}} \le \mathrm{Mass} + \mathrm{M} \cdot (1 - \delta_{\mathrm{atm}}(i)), \quad i \in [n_{\mathrm{LB}} + \mathrm{na}_{\mathrm{LB}}(\mathtt{H}), n^* + \mathrm{na}_{\mathrm{UB}}(\mathtt{H})]. \tag{64}$$

5.7 Constraints for Bounds on the Number of Bonds

We include constraints for specification of lower and upper bounds $\mathrm{bd}_{\mathrm{LB}}$ and $\mathrm{bd}_{\mathrm{UB}}$.

constants:

- $\mathrm{bd}_{m,\mathrm{LB}}(i), \mathrm{bd}_{m,\mathrm{UB}}(i) \in [0, \mathrm{n}^{\mathrm{int}}_{\mathrm{UB}}]$, $i \in [1, m_{\mathrm{C}}]$, $m \in [2, 3]$: 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 $e_i \in E_{\mathrm{C}}$;

variables :

- $\mathrm{bd}_{\mathrm{T}}(k, i, m) \in [0, 1]$, $k \in [1, k_{\mathrm{C}}]$, $i \in [2, t_{\mathrm{T}}]$, $m \in [2, 3]$: $\mathrm{bd}_{\mathrm{T}}(k, i, m) = 1 \Leftrightarrow$ the pure path P_k for edge $e_k \in E_{\mathrm{C}}$ contains edge ${e^{\mathrm{T}}}_i$ with $\beta({e^{\mathrm{T}}}_i) = m$;

constraints:

$$\mathrm{bd}_{m,\mathrm{LB}}(i) \le \delta^{\mathrm{C}}_{\beta}(i, m) \le \mathrm{bd}_{m,\mathrm{UB}}(i), i \in I_{(=1)} \cup I_{(0/1)}, m \in [2, 3], \tag{65}$$

$$\mathrm{bd}_{\mathrm{T}}(k, i, m) \ge \delta^{\mathrm{T}}_{\beta}(i, m) + \chi^{\mathrm{T}}(i, k) - 1, \quad k \in [1, k_{\mathrm{C}}], i \in [2, t_{\mathrm{T}}], m \in [2, 3], \tag{66}$$

$$\sum_{j\in[2,t_{\mathrm{T}}]} \delta^{\mathrm{T}}_{\beta}(j, m) \ge \sum_{k\in[1,k_{\mathrm{C}}], i\in[2,t_{\mathrm{T}}]} \mathrm{bd}_{\mathrm{T}}(k, i, m), \quad m \in [2, 3], \tag{67}$$

$$\mathrm{bd}_{m,\mathrm{LB}}(k) \le \sum_{i\in[2,t_{\mathrm{T}}]} \mathrm{bd}_{\mathrm{T}}(k, i, m) + \delta^{\mathrm{CT}}_{\beta}(k, m) + \delta^{\mathrm{TC}}_{\beta}(k, m) \le \mathrm{bd}_{m,\mathrm{UB}}(k), \quad k \in [1, k_{\mathrm{C}}], m \in [2, 3]. \tag{68}$$

5.8 Descriptor for the Number of Adjacency-configurations

We call a tuple $(\mathtt{a}, \mathtt{b}, m) \in (\Lambda \setminus \{\mathtt{H}\}) \times (\Lambda \setminus \{\mathtt{H}\}) \times [1, 3]$ an *adjacency-configuration*. The adjacency-configuration of an edge-configuration $(\mu = \mathtt{a}d, \mu' = \mathtt{b}d', m)$ is defined to be $(\mathtt{a}, \mathtt{b}, m)$. We include constraints to compute the frequency of each adjacency-configuration in an inferred chemical graph $\mathbb{C}$.

constants:

- A set Γ^{int} of edge-configurations $\gamma = (\mu, \mu', m)$ with $\mu \leq \mu'$;
- Let $\overline{\gamma}$ of an edge-configuration $\gamma = (\mu, \mu', m)$ denote the edge-configuration (μ', μ, m);
- Let $\Gamma^{\mathrm{int}}_{<} = \{(\mu, \mu', m) \in \Gamma^{\mathrm{int}} \mid \mu < \mu'\}$, $\Gamma^{\mathrm{int}}_{=} = \{(\mu, \mu', m) \in \Gamma^{\mathrm{int}} \mid \mu = \mu'\}$ and $\Gamma^{\mathrm{int}}_{>} = \{\overline{\gamma} \mid \gamma \in \Gamma^{\mathrm{int}}_{<}\}$;
- Let $\Gamma^{\mathrm{int}}_{\mathrm{ac},<}$, $\Gamma^{\mathrm{int}}_{\mathrm{ac},=}$ and $\Gamma^{\mathrm{int}}_{\mathrm{ac},>}$ denote the sets of the adjacency-configurations of edge-configurations in the sets $\Gamma^{\mathrm{int}}_{<}$, $\Gamma^{\mathrm{int}}_{=}$ and $\Gamma^{\mathrm{int}}_{>}$, respectively;
- Let $\overline{\nu}$ of an adjacency-configuration $\nu = (\mathtt{a}, \mathtt{b}, m)$ denote the adjacency-configuration $(\mathtt{b}, \mathtt{a}, m)$;
- Prepare a coding of the set $\Gamma^{\mathrm{int}}_{\mathrm{ac}} \cup \Gamma^{\mathrm{int}}_{\mathrm{ac},>}$ and let $[\nu]^{\mathrm{int}}$ denote the coded integer of an element ν in $\Gamma^{\mathrm{int}}_{\mathrm{ac}} \cup \Gamma^{\mathrm{int}}_{\mathrm{ac},>}$;
- Choose subsets $\widetilde{\Gamma}^{\mathrm{C}}_{\mathrm{ac}}, \widetilde{\Gamma}^{\mathrm{T}}_{\mathrm{ac}}, \widetilde{\Gamma}^{\mathrm{CT}}_{\mathrm{ac}}, \widetilde{\Gamma}^{\mathrm{TC}}_{\mathrm{ac}}, \widetilde{\Gamma}^{\mathrm{F}}_{\mathrm{ac}}, \widetilde{\Gamma}^{\mathrm{CF}}_{\mathrm{ac}}, \widetilde{\Gamma}^{\mathrm{TF}}_{\mathrm{ac}} \subseteq \Gamma^{\mathrm{int}}_{\mathrm{ac}} \cup \Gamma^{\mathrm{int}}_{\mathrm{ac},>}$; To compute the frequency of adjacency-configurations exactly, set $\widetilde{\Gamma}^{\mathrm{C}}_{\mathrm{ac}} := \widetilde{\Gamma}^{\mathrm{T}}_{\mathrm{ac}} := \widetilde{\Gamma}^{\mathrm{CT}}_{\mathrm{ac}} := \widetilde{\Gamma}^{\mathrm{TC}}_{\mathrm{ac}} := \widetilde{\Gamma}^{\mathrm{F}}_{\mathrm{ac}} := \widetilde{\Gamma}^{\mathrm{CF}}_{\mathrm{ac}} := \widetilde{\Gamma}^{\mathrm{TF}}_{\mathrm{ac}} := \Gamma^{\mathrm{int}}_{\mathrm{ac}} \cup \Gamma^{\mathrm{int}}_{\mathrm{ac},>}$;
- $\mathrm{ac}^{\mathrm{int}}_{\mathrm{LB}}(\nu), \mathrm{ac}^{\mathrm{int}}_{\mathrm{UB}}(\nu) \in [0, 2\mathrm{n}^{\mathrm{int}}_{\mathrm{UB}}], \nu = (\mathtt{a}, \mathtt{b}, m) \in \Gamma^{\mathrm{int}}_{\mathrm{ac}}$: lower and upper bounds on the number of interior-edges $e = uv$ with $\alpha(u) = \mathtt{a}$, $\alpha(v) = \mathtt{b}$ and $\beta(e) = m$;
- A subset $\Gamma^{\mathrm{lnk}}_{\mathrm{ac}} \subseteq \Gamma^{\mathrm{int}}_{\mathrm{ac}}$ for adjacency-configurations of link-edges. Let $\Gamma^{\mathrm{lnk}}_{\mathrm{ac},<} = \Gamma^{\mathrm{lnk}}_{\mathrm{ac}} \cap \Gamma^{\mathrm{int}}_{\mathrm{ac},<}$, $\Gamma^{\mathrm{lnk}}_{\mathrm{ac},=} = \Gamma^{\mathrm{lnk}}_{\mathrm{ac}} \cap \Gamma^{\mathrm{int}}_{\mathrm{ac},=}$ and $\Gamma^{\mathrm{int}}_{\mathrm{ac},>} = \{(\mathtt{b}, \mathtt{a}, m) \mid (\mathtt{a}, \mathtt{b}, m) \in \Gamma^{\mathrm{lnk}}_{\mathrm{ac},<}\}$;
- $\mathrm{ac}^{\mathrm{lnk}}_{\mathrm{LB}}(\nu), \mathrm{ac}^{\mathrm{lnk}}_{\mathrm{UB}}(\nu) \in [0, 2\mathrm{n}^{\mathrm{int}}_{\mathrm{UB}}], \nu = (\mathtt{a}, \mathtt{b}, m) \in \Gamma^{\mathrm{lnk}}_{\mathrm{ac}}$: lower and upper bounds on the number of link-edges $e = uv$ with $\alpha(u) = \mathtt{a}$, $\alpha(v) = \mathtt{b}$ and $\beta(e) = m$;

variables:

- $\mathrm{ac}^{\mathrm{int}}([\nu]^{\mathrm{int}}) \in [\mathrm{ac}^{\mathrm{int}}_{\mathrm{LB}}(\nu), \mathrm{ac}^{\mathrm{int}}_{\mathrm{UB}}(\nu)], \nu \in \Gamma^{\mathrm{int}}_{\mathrm{ac}}$: the number of interior-edges with adjacency-configuration ν;
- $\mathrm{ac}_{\mathrm{C}}([\nu]^{\mathrm{int}}) \in [0, m_{\mathrm{C}}], \nu \in \widetilde{\Gamma}^{\mathrm{C}}_{\mathrm{ac}}$, $\mathrm{ac}_{\mathrm{T}}([\nu]^{\mathrm{int}}) \in [0, t_{\mathrm{T}}], \nu \in \widetilde{\Gamma}^{\mathrm{T}}_{\mathrm{ac}}$, $\mathrm{ac}_{\mathrm{F}}([\nu]^{\mathrm{int}}) \in [0, t_{\mathrm{F}}], \nu \in \widetilde{\Gamma}^{\mathrm{F}}_{\mathrm{ac}}$: the number of edges $e^{\mathrm{C}} \in E_{\mathrm{C}}$ (resp., edges $e^{\mathrm{T}} \in E_{\mathrm{T}}$ and edges $e^{\mathrm{F}} \in E_{\mathrm{F}}$) with adjacency-configuration ν;
- $\mathrm{ac}_{\mathrm{CT}}([\nu]^{\mathrm{int}}) \in [0, \min\{k_{\mathrm{C}}, t_{\mathrm{T}}\}], \nu \in \widetilde{\Gamma}^{\mathrm{CT}}_{\mathrm{ac}}$, $\mathrm{ac}_{\mathrm{TC}}([\nu]^{\mathrm{int}}) \in [0, \min\{k_{\mathrm{C}}, t_{\mathrm{T}}\}], \nu \in \widetilde{\Gamma}^{\mathrm{CT}}_{\mathrm{ac}}$, $\mathrm{ac}_{\mathrm{CF}}([\nu]^{\mathrm{int}}) \in [0, \widetilde{t_{\mathrm{C}}}], \nu \in \widetilde{\Gamma}^{\mathrm{CF}}_{\mathrm{ac}}$, $\mathrm{ac}_{\mathrm{TF}}([\nu]^{\mathrm{int}}) \in [0, t_{\mathrm{T}}], \nu \in \widetilde{\Gamma}^{\mathrm{TF}}_{\mathrm{ac}}$: the number of edges $e^{\mathrm{CT}} \in E_{\mathrm{CT}}$ (resp., edges $e^{\mathrm{TC}} \in E_{\mathrm{TC}}$ and edges $e^{\mathrm{CF}} \in E_{\mathrm{CF}}$ and $e^{\mathrm{TF}} \in E_{\mathrm{TF}}$) with adjacency-configuration ν;
- $\delta^{\mathrm{C}}_{\mathrm{ac}}(i, [\nu]^{\mathrm{int}}) \in [0, 1], i \in [\widetilde{k_{\mathrm{C}}} + 1, m_{\mathrm{C}}] = I_{(\geq 1)} \cup I_{(0/1)} \cup I_{(=1)}, \nu \in \widetilde{\Gamma}^{\mathrm{C}}_{\mathrm{ac}}$, $\delta^{\mathrm{T}}_{\mathrm{ac}}(i, [\nu]^{\mathrm{int}}) \in [0, 1], i \in [2, t_{\mathrm{T}}], \nu \in \widetilde{\Gamma}^{\mathrm{T}}_{\mathrm{ac}}$, $\delta^{\mathrm{F}}_{\mathrm{ac}}(i, [\nu]^{\mathrm{int}}) \in [0, 1], i \in [2, t_{\mathrm{F}}], \nu \in \widetilde{\Gamma}^{\mathrm{F}}_{\mathrm{ac}}$: $\delta^{\mathrm{X}}_{\mathrm{ac}}(i, [\nu]^{\mathrm{int}}) = 1 \Leftrightarrow$ edge ${e^{\mathrm{X}}}_i$ has adjacency-configuration ν;
- $\delta^{\mathrm{CT}}_{\mathrm{ac}}(k, [\nu]^{\mathrm{int}}), \delta^{\mathrm{TC}}_{\mathrm{ac}}(k, [\nu]^{\mathrm{int}}) \in [0, 1], k \in [1, k_{\mathrm{C}}] = I_{(\geq 2)} \cup I_{(\geq 1)}, \nu \in \widetilde{\Gamma}^{\mathrm{CT}}_{\mathrm{ac}}$: $\delta^{\mathrm{CT}}_{\mathrm{ac}}(k, [\nu]^{\mathrm{int}}) = 1$ (resp., $\delta^{\mathrm{TC}}_{\mathrm{ac}}(k, [\nu]^{\mathrm{int}}) = 1$) $\Leftrightarrow$ edge ${e^{\mathrm{CT}}}_{\mathrm{tail}(k),j}$ (resp., ${e^{\mathrm{TC}}}_{\mathrm{head}(k),j}$) for some $j \in [1, t_{\mathrm{T}}]$ has adjacency-configuration ν;

- $\delta_{\rm ac}^{\rm CF}(c,[\nu]^{\rm int}) \in [0,1], c \in [1,\widetilde{t_{\rm C}}], \nu \in \widetilde{\Gamma}_{\rm ac}^{\rm CF}$: $\delta_{\rm ac}^{\rm CF}(c,[\nu]^{\rm int}) = 1 \Leftrightarrow$ edge ${e^{\rm CF}}_{c,i}$ for some $i \in [1,t_{\rm F}]$ has adjacency-configuration ν;
- $\delta_{\rm ac}^{\rm TF}(i,[\nu]^{\rm int}) \in [0,1], i \in [1,t_{\rm T}], \nu \in \widetilde{\Gamma}_{\rm ac}^{\rm TF}$: $\delta_{\rm ac}^{\rm TF}(i,[\nu]^{\rm int}) = 1 \Leftrightarrow$ edge ${e^{\rm TF}}_{i,j}$ for some $j \in [1,t_{\rm F}]$ has adjacency-configuration ν;
- $\alpha^{\rm CT}(k), \alpha^{\rm TC}(k) \in [0,|\Lambda^{\rm int}|], k \in [1,k_{\rm C}]$: $\alpha(v)$ of the edge $({v^{\rm C}}_{{\rm tail}(k)}, v) \in E_{\rm CT}$ (resp., $(v,{v^{\rm C}}_{{\rm head}(k)}) \in E_{\rm TC}$) if any;
- $\alpha^{\rm CF}(c) \in [0,|\Lambda^{\rm int}|], c \in [1,\widetilde{t_{\rm C}}]$: $\alpha(v)$ of the edge $({v^{\rm C}}_{c}, v) \in E_{\rm CF}$ if any;
- $\alpha^{\rm TF}(i) \in [0,|\Lambda^{\rm int}|], i \in [1,t_{\rm T}]$: $\alpha(v)$ of the edge $({v^{\rm T}}_{i}, v) \in E_{\rm TF}$ if any;
- $\Delta_{\rm ac}^{\rm C+}(i), \Delta_{\rm ac}^{\rm C-}(i), \in [0,|\Lambda^{\rm int}|], i \in [\widetilde{k_{\rm C}}{+}1, m_{\rm C}]$, $\Delta_{\rm ac}^{\rm T+}(i), \Delta_{\rm ac}^{\rm T-}(i) \in [0,|\Lambda^{\rm int}|], i \in [2,t_{\rm T}]$, $\Delta_{\rm ac}^{\rm F+}(i), \Delta_{\rm ac}^{\rm F-}(i) \in [0,|\Lambda^{\rm int}|], i \in [2,t_{\rm F}]$: $\Delta_{\rm ac}^{\rm X+}(i) = \Delta_{\rm ac}^{\rm X-}(i) = 0$ (resp., $\Delta_{\rm ac}^{\rm X+}(i) = \alpha(u)$ and $\Delta_{\rm ac}^{\rm X-}(i) = \alpha(v)$) $\Leftrightarrow$ edge ${e^{\rm X}}_i = (u,v) \in E_{\rm X}$ is used in $\mathbb{C}$ (resp., ${e^{\rm X}}_i \notin E(G)$);
- $\Delta_{\rm ac}^{\rm CT+}(k), \Delta_{\rm ac}^{\rm CT-}(k) \in [0,|\Lambda^{\rm int}|], k \in [1,k_{\rm C}] = I_{(\geq 2)} \cup I_{(\geq 1)}$: $\Delta_{\rm ac}^{\rm CT+}(k) = \Delta_{\rm ac}^{\rm CT-}(k) = 0$ (resp., $\Delta_{\rm ac}^{\rm CT+}(k) = \alpha(u)$ and $\Delta_{\rm ac}^{\rm CT-}(k) = \alpha(v)$) $\Leftrightarrow$ edge ${e^{\rm CT}}_{{\rm tail}(k),j} = (u,v) \in E_{\rm CT}$ for some $j \in [1,t_{\rm T}]$ is used in $\mathbb{C}$ (resp., otherwise);
- $\Delta_{\rm ac}^{\rm TC+}(k), \Delta_{\rm ac}^{\rm TC-}(k) \in [0,|\Lambda^{\rm int}|], k \in [1,k_{\rm C}] = I_{(\geq 2)} \cup I_{(\geq 1)}$: Analogous with $\Delta_{\rm ac}^{\rm CT+}(k)$ and $\Delta_{\rm ac}^{\rm CT-}(k)$;
- $\Delta_{\rm ac}^{\rm CF+}(c) \in [0,|\Lambda^{\rm int}|], \Delta_{\rm ac}^{\rm CF-}(c) \in [0,|\Lambda^{\rm int}|], c \in [1,\widetilde{t_{\rm C}}]$: $\Delta_{\rm ac}^{\rm CF+}(c) = \Delta_{\rm ac}^{\rm CF-}(c) = 0$ (resp., $\Delta_{\rm ac}^{\rm CF+}(c) = \alpha(u)$ and $\Delta_{\rm ac}^{\rm CF-}(c) = \alpha(v)$) $\Leftrightarrow$ edge ${e^{\rm CF}}_{c,i} = (u,v) \in E_{\rm CF}$ for some $i \in [1,t_{\rm F}]$ is used in $\mathbb{C}$ (resp., otherwise);
- $\Delta_{\rm ac}^{\rm TF+}(i) \in [0,|\Lambda^{\rm int}|], \Delta_{\rm ac}^{\rm TF-}(i) \in [0,|\Lambda^{\rm int}|], i \in [1,t_{\rm T}]$: Analogous with $\Delta_{\rm ac}^{\rm CF+}(c)$ and $\Delta_{\rm ac}^{\rm CF-}(c)$;
- ${\rm ac}^{\rm lnk}([\nu]^{\rm int}) \in [{\rm ac}_{\rm LB}^{\rm lnk}(\nu), {\rm ac}_{\rm UB}^{\rm lnk}(\nu)], \nu \in \Gamma_{\rm ac}^{\rm lnk}$: the number of link-edges with adjacency-configuration ν;
- ${\rm ac}_{\rm C}^{\rm lnk}([\nu]^{\rm int}), {\rm ac}_{\rm T}^{\rm lnk}([\nu]^{\rm int}) \in [0,m_{\rm C}], \nu \in \Gamma_{\rm ac}^{\rm lnk}$: the number of link-edges $e^{\rm C} \in E_{\rm C}$ (resp., edges $e^{\rm T} \in E_{\rm T}$) with adjacency-configuration ν;
- ${\rm ac}_{\rm CT}^{\rm lnk}([\nu]^{\rm int}) \in [0,\min\{k_{\rm C},t_{\rm T}\}]$, ${\rm ac}_{\rm TC}^{\rm lnk}([\nu]^{\rm int}) \in [0,\min\{k_{\rm C},t_{\rm T}\}], \nu \in \Gamma_{\rm ac}^{\rm lnk}$: the number of link-edges $e^{\rm CT} \in E_{\rm CT}$ (resp., link-edges $e^{\rm TC} \in E_{\rm TC}$) with adjacency-configuration ν;
- $\delta_{\rm ac}^{\rm T,lnk}(i,[\nu]^{\rm int}) \in [0,1], i \in [2,t_{\rm T}], \nu \in \Gamma_{\rm ac}^{\rm lnk}$: $\delta_{\rm ac}^{\rm T,lnk}(i,[\nu]^{\rm int}) = 1 \Leftrightarrow$ edge ${e^{\rm T}}_i \in E_{\rm T}$ is a link-edge with adjacency-configuration ν;

constraints:

$$
\begin{aligned}
{\rm ac}_{\rm C}([\nu]^{\rm int}) &= 0, & \nu &\in \Gamma_{\rm ac}^{\rm int} \setminus \widetilde{\Gamma}_{\rm ac}^{\rm C}, \\
{\rm ac}_{\rm T}([\nu]^{\rm int}) &= 0, & \nu &\in \Gamma_{\rm ac}^{\rm int} \setminus \widetilde{\Gamma}_{\rm ac}^{\rm T}, \\
{\rm ac}_{\rm F}([\nu]^{\rm int}) &= 0, & \nu &\in \Gamma_{\rm ac}^{\rm int} \setminus \widetilde{\Gamma}_{\rm ac}^{\rm F}, \\
{\rm ac}_{\rm CT}([\nu]^{\rm int}) &= 0, & \nu &\in \Gamma_{\rm ac}^{\rm int} \setminus \widetilde{\Gamma}_{\rm ac}^{\rm CT}, \\
{\rm ac}_{\rm TC}([\nu]^{\rm int}) &= 0, & \nu &\in \Gamma_{\rm ac}^{\rm int} \setminus \widetilde{\Gamma}_{\rm ac}^{\rm TC}, \\
{\rm ac}_{\rm CF}([\nu]^{\rm int}) &= 0, & \nu &\in \Gamma_{\rm ac}^{\rm int} \setminus \widetilde{\Gamma}_{\rm ac}^{\rm CF}, \\
{\rm ac}_{\rm TF}([\nu]^{\rm int}) &= 0, & \nu &\in \Gamma_{\rm ac}^{\rm int} \setminus \widetilde{\Gamma}_{\rm ac}^{\rm TF},
\end{aligned}
\tag{69}
$$

$$\begin{aligned}
\sum_{(\mathsf{a},\mathsf{b},m)=\nu\in\Gamma_{\mathrm{ac}}^{\mathrm{int}}} \mathrm{ac}_{\mathrm{C}}([\nu]^{\mathrm{int}}) &= \sum_{i\in[\widetilde{k_{\mathrm{C}}}+1,m_{\mathrm{C}}]} \delta_\beta^{\mathrm{C}}(i,m), && m\in[1,3],\\
\sum_{(\mathsf{a},\mathsf{b},m)=\nu\in\Gamma_{\mathrm{ac}}^{\mathrm{int}}} \mathrm{ac}_{\mathrm{T}}([\nu]^{\mathrm{int}}) &= \sum_{i\in[2,t_{\mathrm{T}}]} \delta_\beta^{\mathrm{T}}(i,m), && m\in[1,3],\\
\sum_{(\mathsf{a},\mathsf{b},m)=\nu\in\Gamma_{\mathrm{ac}}^{\mathrm{int}}} \mathrm{ac}_{\mathrm{F}}([\nu]^{\mathrm{int}}) &= \sum_{i\in[2,t_{\mathrm{F}}]} \delta_\beta^{\mathrm{F}}(i,m), && m\in[1,3],\\
\sum_{(\mathsf{a},\mathsf{b},m)=\nu\in\Gamma_{\mathrm{ac}}^{\mathrm{int}}} \mathrm{ac}_{\mathrm{CT}}([\nu]^{\mathrm{int}}) &= \sum_{k\in[1,k_{\mathrm{C}}]} \delta_\beta^{\mathrm{CT}}(k,m), && m\in[1,3],\\
\sum_{(\mathsf{a},\mathsf{b},m)=\nu\in\Gamma_{\mathrm{ac}}^{\mathrm{int}}} \mathrm{ac}_{\mathrm{TC}}([\nu]^{\mathrm{int}}) &= \sum_{k\in[1,k_{\mathrm{C}}]} \delta_\beta^{\mathrm{TC}}(k,m), && m\in[1,3],\\
\sum_{(\mathsf{a},\mathsf{b},m)=\nu\in\Gamma_{\mathrm{ac}}^{\mathrm{int}}} \mathrm{ac}_{\mathrm{CF}}([\nu]^{\mathrm{int}}) &= \sum_{c\in[1,\widetilde{t_{\mathrm{C}}}]} \delta_\beta^{*\mathrm{F}}(c,m), && m\in[1,3],\\
\sum_{(\mathsf{a},\mathsf{b},m)=\nu\in\Gamma_{\mathrm{ac}}^{\mathrm{int}}} \mathrm{ac}_{\mathrm{TF}}([\nu]^{\mathrm{int}}) &= \sum_{c\in[\widetilde{t_{\mathrm{C}}}+1,c_{\mathrm{F}}]} \delta_\beta^{*\mathrm{F}}(c,m), && m\in[1,3],
\end{aligned} \tag{70}$$

$$\begin{aligned}
\sum_{\nu=(\mathsf{a},\mathsf{b},m)\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{C}}} m\cdot\delta_{\mathrm{ac}}^{\mathrm{C}}(i,[\nu]^{\mathrm{int}}) &= \beta^{\mathrm{C}}(i),\\
\Delta_{\mathrm{ac}}^{\mathrm{C}+}(i) + \sum_{\nu=(\mathsf{a},\mathsf{b},m)\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{C}}} [\mathsf{a}]^{\mathrm{int}}\delta_{\mathrm{ac}}^{\mathrm{C}}(i,[\nu]^{\mathrm{int}}) &= \alpha^{\mathrm{C}}(\mathrm{tail}(i)),\\
\Delta_{\mathrm{ac}}^{\mathrm{C}-}(i) + \sum_{\nu=(\mathsf{a},\mathsf{b},m)\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{C}}} [\mathsf{b}]^{\mathrm{int}}\delta_{\mathrm{ac}}^{\mathrm{C}}(i,[\nu]^{\mathrm{int}}) &= \alpha^{\mathrm{C}}(\mathrm{head}(i)),\\
\Delta_{\mathrm{ac}}^{\mathrm{C}+}(i) + \Delta_{\mathrm{ac}}^{\mathrm{C}-}(i) &\le 2|\Lambda^{\mathrm{int}}|(1-e^{\mathrm{C}}(i)), && i\in[\widetilde{k_{\mathrm{C}}}+1,m_{\mathrm{C}}],\\
\sum_{i\in[\widetilde{k_{\mathrm{C}}}+1,m_{\mathrm{C}}]} \delta_{\mathrm{ac}}^{\mathrm{C}}(i,[\nu]^{\mathrm{int}}) &= \mathrm{ac}_{\mathrm{C}}([\nu]^{\mathrm{int}}), && \nu\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{C}},
\end{aligned} \tag{71}$$

$$\sum_{i\in I_{\mathrm{lnk}}\cap[\widetilde{k_{\mathrm{C}}}+1,m_{\mathrm{C}}]} \delta_{\mathrm{ac}}^{\mathrm{C}}(i,[\nu]^{\mathrm{int}}) = \mathrm{ac}_{\mathrm{C}}^{\mathrm{lnk}}([\nu]^{\mathrm{int}}), \qquad \nu\in\Gamma_{\mathrm{ac}}^{\mathrm{lnk}}\cup\Gamma_{\mathrm{ac},>}^{\mathrm{lnk}}, \tag{72}$$

$$\begin{aligned}
\sum_{\nu=(\mathsf{a},\mathsf{b},m)\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{T}}} m\cdot\delta_{\mathrm{ac}}^{\mathrm{T}}(i,[\nu]^{\mathrm{int}}) &= \beta^{\mathrm{T}}(i),\\
\Delta_{\mathrm{ac}}^{\mathrm{T}+}(i) + \sum_{\nu=(\mathsf{a},\mathsf{b},m)\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{T}}} [\mathsf{a}]^{\mathrm{int}}\delta_{\mathrm{ac}}^{\mathrm{T}}(i,[\nu]^{\mathrm{int}}) &= \alpha^{\mathrm{T}}(i-1),\\
\Delta_{\mathrm{ac}}^{\mathrm{T}-}(i) + \sum_{\nu=(\mathsf{a},\mathsf{b},m)\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{T}}} [\mathsf{b}]^{\mathrm{int}}\delta_{\mathrm{ac}}^{\mathrm{T}}(i,[\nu]^{\mathrm{int}}) &= \alpha^{\mathrm{T}}(i),\\
\Delta_{\mathrm{ac}}^{\mathrm{T}+}(i) + \Delta_{\mathrm{ac}}^{\mathrm{T}-}(i) &\le 2|\Lambda^{\mathrm{int}}|(1-e^{\mathrm{T}}(i)), && i\in[2,t_{\mathrm{T}}],\\
\sum_{i\in[2,t_{\mathrm{T}}]} \delta_{\mathrm{ac}}^{\mathrm{T}}(i,[\nu]^{\mathrm{int}}) &= \mathrm{ac}_{\mathrm{T}}([\nu]^{\mathrm{int}}), && \nu\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{T}},
\end{aligned} \tag{73}$$

$$\begin{aligned}
\delta^{\mathrm{T}}_{\mathrm{ac}}(i,[\nu]^{\mathrm{int}})+\sum_{k\in I_{\mathrm{lnk}}\cap[1,k_{\mathrm{C}}]}\chi^{\mathrm{T}}(i,k) &\geq 2\delta^{\mathrm{T,lnk}}_{\mathrm{ac}}(i,[\nu]^{\mathrm{int}}), \quad i\in[2,t_{\mathrm{T}}],\\
\delta^{\mathrm{T,lnk}}_{\mathrm{ac}}(i,[\nu]^{\mathrm{int}}) &\geq \delta^{\mathrm{T}}_{\mathrm{ac}}(i,[\nu]^{\mathrm{int}})+\sum_{k\in I_{\mathrm{lnk}}\cap[1,k_{\mathrm{C}}]}\chi^{\mathrm{T}}(i,k)-1, \quad i\in[2,t_{\mathrm{T}}],\\
\sum_{i\in[2,t_{\mathrm{T}}]}\delta^{\mathrm{T,lnk}}_{\mathrm{ac}}(i,[\nu]^{\mathrm{int}}) &= \mathrm{ac}^{\mathrm{lnk}}_{\mathrm{T}}([\nu]^{\mathrm{int}}), \qquad \nu\in\Gamma^{\mathrm{lnk}}_{\mathrm{ac}}\cup\Gamma^{\mathrm{lnk}}_{\mathrm{ac},>},
\end{aligned}\tag{74}$$

$$\begin{aligned}
\sum_{\nu=(\mathtt{a},\mathtt{b},m)\in\widetilde{\Gamma}^{\mathrm{F}}_{\mathrm{ac}}} m\cdot\delta^{\mathrm{F}}_{\mathrm{ac}}(i,[\nu]^{\mathrm{int}}) &= \beta^{\mathrm{F}}(i),\\
\Delta^{\mathrm{F}+}_{\mathrm{ac}}(i)+\sum_{\nu=(\mathtt{a},\mathtt{b},m)\in\widetilde{\Gamma}^{\mathrm{F}}_{\mathrm{ac}}}[\mathtt{a}]^{\mathrm{int}}\delta^{\mathrm{F}}_{\mathrm{ac}}(i,[\nu]^{\mathrm{int}}) &= \alpha^{\mathrm{F}}(i-1),\\
\Delta^{\mathrm{F}-}_{\mathrm{ac}}(i)+\sum_{\nu=(\mathtt{a},\mathtt{b},m)\in\widetilde{\Gamma}^{\mathrm{F}}_{\mathrm{ac}}}[\mathtt{b}]^{\mathrm{int}}\delta^{\mathrm{F}}_{\mathrm{ac}}(i,[\nu]^{\mathrm{int}}) &= \alpha^{\mathrm{F}}(i),\\
\Delta^{\mathrm{F}+}_{\mathrm{ac}}(i)+\Delta^{\mathrm{F}-}_{\mathrm{ac}}(i) &\leq 2|\Lambda^{\mathrm{ex}}|(1-e^{\mathrm{F}}(i)), \qquad i\in[2,t_{\mathrm{F}}],\\
\sum_{i\in[2,t_{\mathrm{F}}]}\delta^{\mathrm{F}}_{\mathrm{ac}}(i,[\nu]^{\mathrm{int}}) &= \mathrm{ac}_{\mathrm{F}}([\nu]^{\mathrm{int}}), \qquad \nu\in\widetilde{\Gamma}^{\mathrm{F}}_{\mathrm{ac}},
\end{aligned}\tag{75}$$

$$\begin{aligned}
\alpha^{\mathrm{T}}(i)+|\Lambda^{\mathrm{int}}|(1-\chi^{\mathrm{T}}(i,k)+e^{\mathrm{T}}(i)) &\geq \alpha^{\mathrm{CT}}(k),\\
\alpha^{\mathrm{CT}}(k) &\geq \alpha^{\mathrm{T}}(i)-|\Lambda^{\mathrm{int}}|(1-\chi^{\mathrm{T}}(i,k)+e^{\mathrm{T}}(i)), \qquad i\in[1,t_{\mathrm{T}}],\\
\sum_{\nu=(\mathtt{a},\mathtt{b},m)\in\widetilde{\Gamma}^{\mathrm{CT}}_{\mathrm{ac}}} m\cdot\delta^{\mathrm{CT}}_{\mathrm{ac}}(k,[\nu]^{\mathrm{int}}) &= \beta^{\mathrm{CT}}(k),\\
\Delta^{\mathrm{CT}+}_{\mathrm{ac}}(k)+\sum_{\nu=(\mathtt{a},\mathtt{b},m)\in\widetilde{\Gamma}^{\mathrm{CT}}_{\mathrm{ac}}}[\mathtt{a}]^{\mathrm{int}}\delta^{\mathrm{CT}}_{\mathrm{ac}}(k,[\nu]^{\mathrm{int}}) &= \alpha^{\mathrm{C}}(\mathrm{tail}(k)),\\
\Delta^{\mathrm{CT}-}_{\mathrm{ac}}(k)+\sum_{\nu=(\mathtt{a},\mathtt{b},m)\in\widetilde{\Gamma}^{\mathrm{CT}}_{\mathrm{ac}}}[\mathtt{b}]^{\mathrm{int}}\delta^{\mathrm{CT}}_{\mathrm{ac}}(k,[\nu]^{\mathrm{int}}) &= \alpha^{\mathrm{CT}}(k),\\
\Delta^{\mathrm{CT}+}_{\mathrm{ac}}(k)+\Delta^{\mathrm{CT}-}_{\mathrm{ac}}(k) &\leq 2|\Lambda^{\mathrm{int}}|(1-\delta^{\mathrm{T}}_{\chi}(k)), \qquad k\in[1,k_{\mathrm{C}}],\\
\sum_{k\in[1,k_{\mathrm{C}}]}\delta^{\mathrm{CT}}_{\mathrm{ac}}(k,[\nu]^{\mathrm{int}}) &= \mathrm{ac}_{\mathrm{CT}}([\nu]^{\mathrm{int}}), \qquad \nu\in\widetilde{\Gamma}^{\mathrm{CT}}_{\mathrm{ac}},
\end{aligned}\tag{76}$$

$$\sum_{i\in I_{\mathrm{lnk}}\cap[1,k_{\mathrm{C}}]}\delta^{\mathrm{CT}}_{\mathrm{ac}}(i,[\nu]^{\mathrm{int}}) = \mathrm{ac}^{\mathrm{lnk}}_{\mathrm{CT}}([\nu]^{\mathrm{int}}), \qquad \nu\in\Gamma^{\mathrm{lnk}}_{\mathrm{ac}}\cup\Gamma^{\mathrm{lnk}}_{\mathrm{ac},>}, \tag{77}$$

$$\begin{aligned}
\alpha^{\mathrm{T}}(i)+|\Lambda^{\mathrm{int}}|(1-\chi^{\mathrm{T}}(i,k)+e^{\mathrm{T}}(i+1)) &\geq \alpha^{\mathrm{TC}}(k), \\
\alpha^{\mathrm{TC}}(k) &\geq \alpha^{\mathrm{T}}(i)-|\Lambda^{\mathrm{int}}|(1-\chi^{\mathrm{T}}(i,k)+e^{\mathrm{T}}(i+1)), && i\in[1,t_{\mathrm{T}}], \\
\sum_{\nu=(\mathtt{a},\mathtt{b},m)\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{TC}}} m\cdot\delta_{\mathrm{ac}}^{\mathrm{TC}}(k,[\nu]^{\mathrm{int}}) &= \beta^{\mathrm{TC}}(k), \\
\Delta_{\mathrm{ac}}^{\mathrm{TC}+}(k)+\sum_{\nu=(\mathtt{a},\mathtt{b},m)\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{TC}}} [\mathtt{a}]^{\mathrm{int}}\delta_{\mathrm{ac}}^{\mathrm{TC}}(k,[\nu]^{\mathrm{int}}) &= \alpha^{\mathrm{TC}}(k), \\
\Delta_{\mathrm{ac}}^{\mathrm{TC}-}(k)+\sum_{\nu=(\mathtt{a},\mathtt{b},m)\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{TC}}} [\mathtt{b}]^{\mathrm{int}}\delta_{\mathrm{ac}}^{\mathrm{TC}}(k,[\nu]^{\mathrm{int}}) &= \alpha^{\mathrm{C}}(\mathrm{head}(k)), \\
\Delta_{\mathrm{ac}}^{\mathrm{TC}+}(k)+\Delta_{\mathrm{ac}}^{\mathrm{TC}-}(k) &\leq 2|\Lambda^{\mathrm{int}}|(1-\delta_{\chi}^{\mathrm{T}}(k)), && k\in[1,k_{\mathrm{C}}], \\
\sum_{k\in[1,k_{\mathrm{C}}]}\delta_{\mathrm{ac}}^{\mathrm{TC}}(k,[\nu]^{\mathrm{int}}) &= \mathrm{ac}_{\mathrm{TC}}([\nu]^{\mathrm{int}}), && \nu\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{TC}},
\end{aligned} \tag{78}$$

$$\sum_{i\in I_{\mathrm{lnk}}\cap[1,k_{\mathrm{C}}]}\delta_{\mathrm{ac}}^{\mathrm{TC}}(i,[\nu]^{\mathrm{int}}) = \mathrm{ac}_{\mathrm{TC}}^{\mathrm{lnk}}([\nu]^{\mathrm{int}}), \qquad \nu\in\Gamma_{\mathrm{ac}}^{\mathrm{lnk}}\cup\Gamma_{\mathrm{ac},>}^{\mathrm{lnk}}, \tag{79}$$

$$\begin{aligned}
\alpha^{\mathrm{F}}(i)+|\Lambda^{\mathrm{int}}|(1-\chi^{\mathrm{F}}(i,c)+e^{\mathrm{F}}(i)) &\geq \alpha^{\mathrm{CF}}(c), \\
\alpha^{\mathrm{CF}}(c) &\geq \alpha^{\mathrm{F}}(i)-|\Lambda^{\mathrm{int}}|(1-\chi^{\mathrm{F}}(i,c)+e^{\mathrm{F}}(i)), && i\in[1,t_{\mathrm{F}}], \\
\sum_{\nu=(\mathtt{a},\mathtt{b},m)\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{CF}}} m\cdot\delta_{\mathrm{ac}}^{\mathrm{CF}}(c,[\nu]^{\mathrm{int}}) &= \beta^{*\mathrm{F}}(c), \\
\Delta_{\mathrm{ac}}^{\mathrm{CF}+}(c)+\sum_{\nu=(\mathtt{a},\mathtt{b},m)\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{CF}}} [\mathtt{a}]^{\mathrm{int}}\delta_{\mathrm{ac}}^{\mathrm{CF}}(c,[\nu]^{\mathrm{int}}) &= \alpha^{\mathrm{C}}(\mathrm{head}(c)), \\
\Delta_{\mathrm{ac}}^{\mathrm{CF}-}(c)+\sum_{\nu=(\mathtt{a},\mathtt{b},m)\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{CF}}} [\mathtt{b}]^{\mathrm{int}}\delta_{\mathrm{ac}}^{\mathrm{CF}}(c,[\nu]^{\mathrm{int}}) &= \alpha^{\mathrm{CF}}(c), \\
\Delta_{\mathrm{ac}}^{\mathrm{CF}+}(c)+\Delta_{\mathrm{ac}}^{\mathrm{CF}-}(c) &\leq 2\max\{|\Lambda^{\mathrm{int}}|,|\Lambda^{\mathrm{int}}|\}(1-\delta_{\chi}^{\mathrm{F}}(c)), && c\in[1,\widetilde{t_{\mathrm{C}}}], \\
\sum_{c\in[1,\widetilde{t_{\mathrm{C}}}]}\delta_{\mathrm{ac}}^{\mathrm{CF}}(c,[\nu]^{\mathrm{int}}) &= \mathrm{ac}_{\mathrm{CF}}([\nu]^{\mathrm{int}}), && \nu\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{CF}},
\end{aligned} \tag{80}$$

$$\begin{aligned}
\alpha^{\mathrm{F}}(j)+|\Lambda^{\mathrm{int}}|(1-\chi^{\mathrm{F}}(j,i+\widetilde{t_{\mathrm{C}}})+e^{\mathrm{F}}(j)) &\geq \alpha^{\mathrm{TF}}(i), \\
\alpha^{\mathrm{TF}}(i) &\geq \alpha^{\mathrm{F}}(j)-|\Lambda^{\mathrm{int}}|(1-\chi^{\mathrm{F}}(j,i+\widetilde{t_{\mathrm{C}}})+e^{\mathrm{F}}(j)), && j\in[1,t_{\mathrm{F}}], \\
\sum_{\nu=(\mathtt{a},\mathtt{b},m)\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{TF}}} m\cdot\delta_{\mathrm{ac}}^{\mathrm{TF}}(i,[\nu]^{\mathrm{int}}) &= \beta^{*\mathrm{F}}(i+\widetilde{t_{\mathrm{C}}}), \\
\Delta_{\mathrm{ac}}^{\mathrm{TF}+}(i)+\sum_{\nu=(\mathtt{a},\mathtt{b},m)\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{TF}}} [\mathtt{a}]^{\mathrm{int}}\delta_{\mathrm{ac}}^{\mathrm{TF}}(i,[\nu]^{\mathrm{int}}) &= \alpha^{\mathrm{T}}(i), \\
\Delta_{\mathrm{ac}}^{\mathrm{TF}-}(i)+\sum_{\nu=(\mathtt{a},\mathtt{b},m)\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{TF}}} [\mathtt{b}]^{\mathrm{int}}\delta_{\mathrm{ac}}^{\mathrm{TF}}(i,[\nu]^{\mathrm{int}}) &= \alpha^{\mathrm{TF}}(i), \\
\Delta_{\mathrm{ac}}^{\mathrm{TF}+}(i)+\Delta_{\mathrm{ac}}^{\mathrm{TF}-}(i) &\leq 2\max\{|\Lambda^{\mathrm{int}}|,|\Lambda^{\mathrm{int}}|\}(1-\delta_{\chi}^{\mathrm{F}}(i+\widetilde{t_{\mathrm{C}}})), && i\in[1,t_{\mathrm{T}}], \\
\sum_{i\in[1,t_{\mathrm{T}}]}\delta_{\mathrm{ac}}^{\mathrm{TF}}(i,[\nu]^{\mathrm{int}}) &= \mathrm{ac}_{\mathrm{TF}}([\nu]^{\mathrm{int}}), && \nu\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{TF}},
\end{aligned} \tag{81}$$

$$\sum_{\mathrm{X}\in\{\mathrm{C,T,F,CT,TC,CF,TF}\}} (\mathrm{ac}_{\mathrm{X}}([\nu]^{\mathrm{int}}) + \mathrm{ac}_{\mathrm{X}}([\overline{\nu}]^{\mathrm{int}})) = \mathrm{ac}^{\mathrm{int}}([\nu]^{\mathrm{int}}), \qquad \nu \in \Gamma^{\mathrm{int}}_{\mathrm{ac},<},$$
$$\sum_{\mathrm{X}\in\{\mathrm{C,T,F,CT,TC,CF,TF}\}} \mathrm{ac}_{\mathrm{X}}([\nu]^{\mathrm{int}}) = \mathrm{ac}^{\mathrm{int}}([\nu]^{\mathrm{int}}), \qquad \nu \in \Gamma^{\mathrm{int}}_{\mathrm{ac},=}, \tag{82}$$

$$\sum_{\mathrm{X}\in\{\mathrm{C,T,CT,TC}\}} (\mathrm{ac}^{\mathrm{lnk}}_{\mathrm{X}}([\nu]^{\mathrm{int}}) + \mathrm{ac}^{\mathrm{lnk}}_{\mathrm{X}}([\overline{\nu}]^{\mathrm{int}})) = \mathrm{ac}^{\mathrm{lnk}}([\nu]^{\mathrm{int}}), \qquad \nu \in \Gamma^{\mathrm{lnk}}_{\mathrm{ac},<},$$
$$\sum_{\mathrm{X}\in\{\mathrm{C,T,CT,TC}\}} \mathrm{ac}^{\mathrm{lnk}}_{\mathrm{X}}([\nu]^{\mathrm{int}}) = \mathrm{ac}^{\mathrm{lnk}}([\nu]^{\mathrm{int}}), \qquad \nu \in \Gamma^{\mathrm{lnk}}_{\mathrm{ac},=}, \tag{83}$$

$$\sum_{\nu\in\nu\in\Gamma^{\mathrm{lnk}}_{\mathrm{ac}}} \mathrm{ac}^{\mathrm{lnk}}([\nu]^{\mathrm{int}}) = n_{\mathrm{lnk}}. \tag{84}$$

5.9 Descriptor for the Number of Chemical Symbols

We include constraints for computing the frequency of each chemical symbol in Λ_{dg}. Let $\mathrm{cs}(v)$ denote the chemical symbol of an interior-vertex v in a chemical graph $\mathbb{C}$ to be inferred; i.e., $\mathrm{cs}(v) = \mu = \mathtt{a}d \in \Lambda_{\mathrm{dg}}$ such that $\alpha(v) = \mathtt{a}$ and $\deg_{\langle\mathbb{C}\rangle}(v) = \deg_H(v) - \deg^{\mathrm{hyd}}_{\mathbb{C}}(v) = d$ in $\mathbb{C} = (H, \alpha, \beta)$.

constants:

- A set $\Lambda^{\mathrm{int}}_{\mathrm{dg}}$ of chemical symbols;
- Prepare a coding of each of the two sets $\Lambda^{\mathrm{int}}_{\mathrm{dg}}$ and let $[\mu]^{\mathrm{int}}$ denote the coded integer of an element $\mu \in \Lambda^{\mathrm{int}}_{\mathrm{dg}}$;
- Choose subsets $\widetilde{\Lambda}^{\mathrm{C}}_{\mathrm{dg}}, \widetilde{\Lambda}^{\mathrm{T}}_{\mathrm{dg}}, \widetilde{\Lambda}^{\mathrm{F}}_{\mathrm{dg}} \subseteq \Lambda^{\mathrm{int}}_{\mathrm{dg}}$: To compute the frequency of chemical symbols exactly, set $\widetilde{\Lambda}^{\mathrm{C}}_{\mathrm{dg}} := \widetilde{\Lambda}^{\mathrm{T}}_{\mathrm{dg}} := \widetilde{\Lambda}^{\mathrm{F}}_{\mathrm{dg}} := \Lambda^{\mathrm{int}}_{\mathrm{dg}}$;

variables:

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

constraints:

$$\sum_{\mu\in\widetilde{\Lambda}^{\mathrm{X}}_{\mathrm{dg}}\cup\{\epsilon\}} \delta^{\mathrm{X}}_{\mathrm{ns}}(i, [\mu]^{\mathrm{int}}) = 1, \qquad \sum_{\mu=\mathtt{a}d\in\widetilde{\Lambda}^{\mathrm{X}}_{\mathrm{dg}}} [\mathtt{a}]^{\mathrm{int}} \cdot \delta^{\mathrm{X}}_{\mathrm{ns}}(i, [\mu]^{\mathrm{int}}) = \alpha^{\mathrm{X}}(i),$$
$$\sum_{\mu=\mathtt{a}d\in\widetilde{\Lambda}^{\mathrm{X}}_{\mathrm{dg}}} d \cdot \delta^{\mathrm{X}}_{\mathrm{ns}}(i, [\mu]^{\mathrm{int}}) = \deg^{\mathrm{X}}(i),$$
$$i \in [1, t_{\mathrm{X}}], \mathrm{X} \in \{\mathrm{C, T, F}\}, \tag{85}$$

$$\sum_{i\in[1,t_{\mathrm{C}}]} \delta^{\mathrm{C}}_{\mathrm{ns}}(i, [\mu]^{\mathrm{int}}) + \sum_{i\in[1,t_{\mathrm{T}}]} \delta^{\mathrm{T}}_{\mathrm{ns}}(i, [\mu]^{\mathrm{int}}) + \sum_{i\in[1,t_{\mathrm{F}}]} \delta^{\mathrm{F}}_{\mathrm{ns}}(i, [\mu]^{\mathrm{int}}) = \mathrm{ns}^{\mathrm{int}}([\mu]^{\mathrm{int}}), \qquad \mu \in \Lambda^{\mathrm{int}}_{\mathrm{dg}}. \tag{86}$$

5.10 Descriptor for the Number of Edge-configurations

We include constraints to compute the frequency of each edge-configuration in an inferred chemical graph $\mathbb{C}$.

constants:

- A set $\Gamma^{\rm int}$ of edge-configurations $\gamma = (\mu, \mu', m)$ with $\mu \le \mu'$, where we let $\overline{\gamma}$ denote (μ', μ, m);
- Let $\Gamma^{\rm int}_{<} = \{(\mu, \mu', m) \in \Gamma^{\rm int} \mid \mu < \mu'\}$, $\Gamma^{\rm int}_{=} = \{(\mu, \mu', m) \in \Gamma^{\rm int} \mid \mu = \mu'\}$ and $\Gamma^{\rm int}_{>} = \{(\mu', \mu, m) \mid (\mu, \mu', m) \in \Gamma^{\rm int}_{<}\}$;
- Prepare a coding of the set $\Gamma^{\rm int} \cup \Gamma^{\rm int}_{>}$ and let $[\gamma]^{\rm int}$ denote the coded integer of an element γ in $\Gamma^{\rm int} \cup \Gamma^{\rm int}_{>}$;
- Choose subsets $\widetilde{\Gamma}^{\rm C}_{\rm ec}, \widetilde{\Gamma}^{\rm T}_{\rm ec}, \widetilde{\Gamma}^{\rm CT}_{\rm ec}, \widetilde{\Gamma}^{\rm TC}_{\rm ec}, \widetilde{\Gamma}^{\rm F}_{\rm ec}, \widetilde{\Gamma}^{\rm CF}_{\rm ec}, \widetilde{\Gamma}^{\rm TF}_{\rm ec} \subseteq \Gamma^{\rm int} \cup \Gamma^{\rm int}_{>}$; To compute the frequency of edge-configurations exactly, set $\widetilde{\Gamma}^{\rm C}_{\rm ec} := \widetilde{\Gamma}^{\rm T}_{\rm ec} := \widetilde{\Gamma}^{\rm CT}_{\rm ec} := \widetilde{\Gamma}^{\rm TC}_{\rm ec} := \widetilde{\Gamma}^{\rm F}_{\rm ec} := \widetilde{\Gamma}^{\rm CF}_{\rm ec} := \widetilde{\Gamma}^{\rm TF}_{\rm ec} := \Gamma^{\rm int} \cup \Gamma^{\rm int}_{>}$;
- $\mathrm{ec}^{\rm int}_{\rm LB}(\gamma), \mathrm{ec}^{\rm int}_{\rm UB}(\gamma) \in [0, 2\mathrm{n}^{\rm int}_{\rm UB}], \gamma = (\mu, \mu', m) \in \Gamma^{\rm int}$: lower and upper bounds on the number of interior-edges $e = uv$ with $\mathrm{cs}(u) = \mu$, $\mathrm{cs}(v) = \mu'$ and $\beta(e) = m$;
- A subset $\Gamma^{\rm lnk} \subseteq \Gamma^{\rm int}$ for edge-configurations of link-edges. Let $\Gamma^{\rm lnk}_{<} = \Gamma^{\rm lnk} \cap \Gamma^{\rm int}_{<}$, $\Gamma^{\rm lnk}_{=} = \Gamma^{\rm lnk} \cap \Gamma^{\rm int}_{=}$ and $\Gamma^{\rm int}_{>} = \{(\mathtt{b}, \mathtt{a}, m) \mid (\mathtt{a}, \mathtt{b}, m) \in \Gamma^{\rm lnk}_{<}\}$;
- $\mathrm{ec}^{\rm lnk}_{\rm LB}(\gamma), \mathrm{ec}^{\rm lnk}_{\rm UB}(\gamma) \in [0, 2\mathrm{n}^{\rm int}_{\rm UB}], \gamma = (\mu, \mu', m) \in \Gamma^{\rm int}$: lower and upper bounds on the number of link-edges $e = uv$ with $\mathrm{cs}(u) = \mu$, $\mathrm{cs}(v) = \mu'$ and $\beta(e) = m$;
- $\mathrm{ns}^{\rm cnt}_{\rm LB}([\mu]), \mathrm{ns}^{\rm cnt}_{\rm UB}([\mu]) \in [0, 2], \mu \in \Lambda^{\rm int}_{\rm dg}$: lower and upper bounds on the number of connecting-vertices v with $\mathrm{cs}(v) = \mu$; Define
  $\Gamma^{\rm cnt}_{<} := \{(\mu, \mu', 1) \in \gamma \in \Gamma^{\rm lnk}_{<} \mid \mu, \mu' \in \Lambda^{\rm int}_{\rm dg}, \mathrm{ns}^{\rm cnt}_{\rm LB}(\mu) \le 1 \le \mathrm{ns}_{\rm UB}(\mu), \mathrm{ns}^{\rm cnt}_{\rm LB}(\mu') \le 1 \le \mathrm{ns}_{\rm UB}(\mu')\}$;
  $\Gamma^{\rm cnt}_{>} := \{(\mu, \mu', 1) \in \gamma \in \Gamma^{\rm lnk}_{>} \mid \mu, \mu' \in \Lambda^{\rm int}_{\rm dg}, \mathrm{ns}^{\rm cnt}_{\rm LB}(\mu) \le 1 \le \mathrm{ns}_{\rm UB}(\mu), \mathrm{ns}^{\rm cnt}_{\rm LB}(\mu') \le 1 \le \mathrm{ns}_{\rm UB}(\mu')\}$;
  $\Gamma^{\rm cnt}_{=} := \{(\mu, \mu, 1) \in \gamma \in \Gamma^{\rm lnk}_{=} \mid \mu \in \Lambda^{\rm int}_{\rm dg}, \mathrm{ns}^{\rm cnt}_{\rm UB}(\mu) = 2\}$;

variables:

- $\mathrm{ec}^{\rm int}([\gamma]^{\rm int}) \in [\mathrm{ec}^{\rm int}_{\rm LB}(\gamma), \mathrm{ec}^{\rm int}_{\rm UB}(\gamma)], \gamma \in \Gamma^{\rm int}$: the number of interior-edges with edge-configuration γ;
- $\mathrm{ec}_{\rm C}([\gamma]^{\rm int}) \in [0, m_{\rm C}], \gamma \in \widetilde{\Gamma}^{\rm C}_{\rm ec}$, $\mathrm{ec}_{\rm T}([\gamma]^{\rm int}) \in [0, t_{\rm T}], \gamma \in \widetilde{\Gamma}^{\rm T}_{\rm ec}$, $\mathrm{ec}_{\rm F}([\gamma]^{\rm int}) \in [0, t_{\rm F}], \gamma \in \widetilde{\Gamma}^{\rm F}_{\rm ec}$: the number of edges $e^{\rm C} \in E_{\rm C}$ (resp., edges $e^{\rm T} \in E_{\rm T}$ and edges $e^{\rm F} \in E_{\rm F}$) with edge-configuration γ;
- $\mathrm{ec}_{\rm CT}([\gamma]^{\rm int}) \in [0, \min\{k_{\rm C}, t_{\rm T}\}], \gamma \in \widetilde{\Gamma}^{\rm CT}_{\rm ec}$, $\mathrm{ec}_{\rm TC}([\gamma]^{\rm int}) \in [0, \min\{k_{\rm C}, t_{\rm T}\}], \gamma \in \widetilde{\Gamma}^{\rm CT}_{\rm ec}$, $\mathrm{ec}_{\rm CF}([\gamma]^{\rm int}) \in [0, \widetilde{t_{\rm C}}], \gamma \in \widetilde{\Gamma}^{\rm CF}_{\rm ec}$, $\mathrm{ec}_{\rm TF}([\gamma]^{\rm int}) \in [0, t_{\rm T}], \gamma \in \widetilde{\Gamma}^{\rm TF}_{\rm ec}$: the number of edges $e^{\rm CT} \in E_{\rm CT}$ (resp., edges $e^{\rm TC} \in E_{\rm TC}$ and edges $e^{\rm CF} \in E_{\rm CF}$ and $e^{\rm TF} \in E_{\rm TF}$) with edge-configuration γ;
- $\delta^{\rm C}_{\rm ec}(i, [\gamma]^{\rm int}) \in [0, 1], i \in [\widetilde{k_{\rm C}} + 1, m_{\rm C}] = I_{(\ge 1)} \cup I_{(0/1)} \cup I_{(=1)}, \gamma \in \widetilde{\Gamma}^{\rm C}_{\rm ec}$, $\delta^{\rm T}_{\rm ec}(i, [\gamma]^{\rm int}) \in [0, 1], i \in [2, t_{\rm T}], \gamma \in \widetilde{\Gamma}^{\rm T}_{\rm ec}$, $\delta^{\rm F}_{\rm ec}(i, [\gamma]^{\rm int}) \in [0, 1], i \in [2, t_{\rm F}], \gamma \in \widetilde{\Gamma}^{\rm F}_{\rm ec}$: $\delta^{\rm X}_{\rm ec}(i, [\gamma]^{\rm t}) = 1 \Leftrightarrow$ edge ${e^{\rm X}}_i$ has edge-configuration γ;
- $\delta^{\rm CT}_{\rm ec,C}(k, [\gamma]^{\rm int}), \delta^{\rm TC}_{\rm ec,C}(k, [\gamma]^{\rm int}) \in [0, 1], k \in [1, k_{\rm C}] = I_{(\ge 2)} \cup I_{(\ge 1)}, \gamma \in \widetilde{\Gamma}^{\rm CT}_{\rm ec}$: $\delta^{\rm CT}_{\rm ec,C}(k, [\gamma]^{\rm int}) = 1$ (resp., $\delta^{\rm TC}_{\rm ec,C}(k, [\gamma]^{\rm int}) = 1$) $\Leftrightarrow$ edge ${e^{\rm CT}}_{{\rm tail}(k),j}$ (resp., ${e^{\rm TC}}_{{\rm head}(k),j}$) for some $j \in [1, t_{\rm T}]$ has edge-configuration γ;
- $\delta^{\rm CF}_{\rm ec,C}(c, [\gamma]^{\rm int}) \in [0, 1], c \in [1, \widetilde{t_{\rm C}}], \gamma \in \widetilde{\Gamma}^{\rm CF}_{\rm ec}$: $\delta^{\rm CF}_{\rm ec,C}(c, [\gamma]^{\rm int}) = 1 \Leftrightarrow$ edge ${e^{\rm CF}}_{c,i}$ for some $i \in [1, t_{\rm F}]$ has edge-configuration γ;

- $\delta^{\rm TF}_{\rm ec,T}(i,[\gamma]^{\rm int})\in[0,1], i\in[1,t_{\rm T}],\gamma\in\widetilde{\Gamma}^{\rm TF}_{\rm ec}$: $\delta^{\rm TF}_{\rm ec,T}(i,[\gamma]^{\rm int})=1\Leftrightarrow$ edge ${e^{\rm TF}}_{i,j}$ for some $j\in[1,t_{\rm F}]$ has edge-configuration γ;
- $\deg^{\rm CT}_{\rm T}(k),\deg^{\rm TC}_{\rm T}(k)\in[0,4], k\in[1,k_{\rm C}]$: $\deg_{\langle\mathbb{C}\rangle}(v)$ of an end-vertex $v\in V_{\rm T}$ of the edge $({v^{\rm C}}_{{\rm tail}(k)},v)\in E_{\rm CT}$ (resp., $(v,{v^{\rm C}}_{{\rm head}(k)})\in E_{\rm TC}$) if any;
- $\deg^{\rm CF}_{\rm F}(c)\in[0,4], c\in[1,\widetilde{t_{\rm C}}]$: $\deg_{\langle\mathbb{C}\rangle}(v)$ of an end-vertex $v\in V_{\rm F}$ of the edge $({v^{\rm C}}_c,v)\in E_{\rm CF}$ if any;
- $\deg^{\rm TF}_{\rm F}(i)\in[0,4], i\in[1,t_{\rm T}]$: $\deg_{\langle\mathbb{C}\rangle}(v)$ of an end-vertex $v\in V_{\rm F}$ of the edge $({v^{\rm T}}_i,v)\in E_{\rm TF}$ if any;
- $\Delta^{\rm C+}_{\rm ec}(i),\Delta^{\rm C-}_{\rm ec}(i),\in[0,4], i\in[\widetilde{k_{\rm C}+1},m_{\rm C}]$, $\Delta^{\rm T+}_{\rm ec}(i),\Delta^{\rm T-}_{\rm ec}(i)\in[0,4], i\in[2,t_{\rm T}]$, $\Delta^{\rm F+}_{\rm ec}(i),\Delta^{\rm F-}_{\rm ec}(i)\in[0,4], i\in[2,t_{\rm F}]$: $\Delta^{\rm X+}_{\rm ec}(i)=\Delta^{\rm X-}_{\rm ec}(i)=0$ (resp., $\Delta^{\rm X+}_{\rm ec}(i)=\deg_{\langle\mathbb{C}\rangle}(u)$ and $\Delta^{\rm X-}_{\rm ec}(i)=\deg_{\langle\mathbb{C}\rangle}(v)$) $\Leftrightarrow$ edge ${e^{\rm X}}_i=(u,v)\in E_{\rm X}$ is used in $\langle\mathbb{C}\rangle$ (resp., ${e^{\rm X}}_i\notin E(\langle\mathbb{C}\rangle)$);
- $\Delta^{\rm CT+}_{\rm ec}(k),\Delta^{\rm CT-}_{\rm ec}(k)\in[0,4], k\in[1,k_{\rm C}]=I_{(\geq2)}\cup I_{(\geq1)}$: $\Delta^{\rm CT+}_{\rm ec}(k)=\Delta^{\rm CT-}_{\rm ec}(k)=0$ (resp., $\Delta^{\rm CT+}_{\rm ec}(k)=\deg_{\langle\mathbb{C}\rangle}(u)$ and $\Delta^{\rm CT-}_{\rm ec}(k)=\deg_{\langle\mathbb{C}\rangle}(v)$) $\Leftrightarrow$ edge ${e^{\rm CT}}_{{\rm tail}(k),j}=(u,v)\in E_{\rm CT}$ for some $j\in[1,t_{\rm T}]$ is used in $\langle\mathbb{C}\rangle$ (resp., otherwise);
- $\Delta^{\rm TC+}_{\rm ec}(k),\Delta^{\rm TC-}_{\rm ec}(k)\in[0,4], k\in[1,k_{\rm C}]=I_{(\geq2)}\cup I_{(\geq1)}$: Analogous with $\Delta^{\rm CT+}_{\rm ec}(k)$ and $\Delta^{\rm CT-}_{\rm ec}(k)$;
- $\Delta^{\rm CF+}_{\rm ac}(c),\Delta^{\rm CF-}_{\rm ec}(c)\in[0,4], c\in[1,\widetilde{t_{\rm C}}]$: $\Delta^{\rm CF+}_{\rm ec}(c)=\Delta^{\rm CF-}_{\rm ec}(c)=0$ (resp., $\Delta^{\rm CF+}_{\rm ec}(c)=\deg_{\langle\mathbb{C}\rangle}(u)$ and $\Delta^{\rm CF-}_{\rm ec}(c)=\deg_{\langle\mathbb{C}\rangle}(v)$) $\Leftrightarrow$ edge ${e^{\rm CF}}_{c,j}=(u,v)\in E_{\rm CF}$ for some $j\in[1,t_{\rm F}]$ is used in $\langle\mathbb{C}\rangle$ (resp., otherwise);
- $\Delta^{\rm TF+}_{\rm ec}(i),\Delta^{\rm TF-}_{\rm ec}(i)\in[0,4], i\in[1,t_{\rm T}]$: Analogous with $\Delta^{\rm CF+}_{\rm ec}(c)$ and $\Delta^{\rm CF-}_{\rm ec}(c)$;
- ${\rm ec}^{\rm lnk}([\gamma]^{\rm int})\in[{\rm ec}^{\rm lnk}_{\rm LB}(\gamma),{\rm ec}^{\rm lnk}_{\rm UB}(\gamma)],\gamma\in\Gamma^{\rm lnk}$: the number of link-edges with edge-configuration γ;
- ${\rm ec}^{\rm lnk}_{\rm C}([\gamma]^{\rm int}),{\rm ec}^{\rm lnk}_{\rm T}([\gamma]^{\rm int})\in[0,m_{\rm C}],\gamma\in\Gamma^{\rm lnk}$: the number of link-edges $e^{\rm C}\in E_{\rm C}$ (resp., edges $e^{\rm T}\in E_{\rm T}$) with edge-configuration γ;
- ${\rm ec}^{\rm lnk}_{\rm CT}([\gamma]^{\rm int})\in[0,\min\{k_{\rm C},t_{\rm T}\}]$, ${\rm ec}^{\rm lnk}_{\rm TC}([\gamma]^{\rm int})\in[0,\min\{k_{\rm C},t_{\rm T}\}],\gamma\in\Gamma^{\rm lnk}$: the number of link-edges $e^{\rm CT}\in E_{\rm CT}$ (resp., link-edges $e^{\rm TC}\in E_{\rm TC}$) with adjacency-configuration γ;
- $\delta^{\rm T,lnk}_{\rm ec}(i,[\gamma]^{\rm int})\in[0,1], i\in[2,t_{\rm T}],\gamma\in\Gamma^{\rm lnk}$: $\delta^{\rm T,lnk}_{\rm ec}(i,[\gamma]^{\rm int})=1\Leftrightarrow$ edge ${e^{\rm T}}_i\in E_{\rm T}$ is a link-edge with edge-configuration γ;
- $\delta^{\rm cnt}([\gamma]^{\rm int})\in[0,1],\gamma\in\Gamma^{\rm cnt}_{<}\cup\Gamma^{\rm cnt}_{=}\cup\Gamma^{\rm cnt}_{>}$: $\delta^{\rm cnt}([\gamma]^{\rm int})=1\Leftrightarrow{\rm ec}(e)=\gamma$ for the link-edge e joining connecting-vertices;

constraints:

$$\begin{aligned}
{\rm ec}_{\rm C}([\gamma]^{\rm int})&=0, && \gamma\in\Gamma^{\rm int}\setminus\widetilde{\Gamma}^{\rm C}_{\rm ec},\\
{\rm ec}_{\rm T}([\gamma]^{\rm int})&=0, && \gamma\in\Gamma^{\rm int}\setminus\widetilde{\Gamma}^{\rm T}_{\rm ec},\\
{\rm ec}_{\rm F}([\gamma]^{\rm int})&=0, && \gamma\in\Gamma^{\rm int}\setminus\widetilde{\Gamma}^{\rm F}_{\rm ec},\\
{\rm ec}_{\rm CT}([\gamma]^{\rm int})&=0, && \gamma\in\Gamma^{\rm int}\setminus\widetilde{\Gamma}^{\rm CT}_{\rm ec},\\
{\rm ec}_{\rm TC}([\gamma]^{\rm int})&=0, && \gamma\in\Gamma^{\rm int}\setminus\widetilde{\Gamma}^{\rm TC}_{\rm ec},\\
{\rm ec}_{\rm CF}([\gamma]^{\rm int})&=0, && \gamma\in\Gamma^{\rm int}\setminus\widetilde{\Gamma}^{\rm CF}_{\rm ec},\\
{\rm ec}_{\rm TF}([\gamma]^{\rm int})&=0, && \gamma\in\Gamma^{\rm int}\setminus\widetilde{\Gamma}^{\rm TF}_{\rm ec},
\end{aligned} \tag{87}$$

$$\begin{aligned}
\sum_{(\mu,\mu',m)=\gamma\in\Gamma^{\mathrm{int}}} \mathrm{ec}_{\mathrm{C}}([\gamma]^{\mathrm{int}}) &= \sum_{i\in[\widetilde{k_{\mathrm{C}}}+1,m_{\mathrm{C}}]} \delta_{\beta}^{\mathrm{C}}(i,m), && m\in[1,3],\\
\sum_{(\mu,\mu',m)=\gamma\in\Gamma^{\mathrm{int}}} \mathrm{ec}_{\mathrm{T}}([\gamma]^{\mathrm{int}}) &= \sum_{i\in[2,t_{\mathrm{T}}]} \delta_{\beta}^{\mathrm{T}}(i,m), && m\in[1,3],\\
\sum_{(\mu,\mu',m)=\gamma\in\Gamma^{\mathrm{int}}} \mathrm{ec}_{\mathrm{F}}([\gamma]^{\mathrm{int}}) &= \sum_{i\in[2,t_{\mathrm{F}}]} \delta_{\beta}^{\mathrm{F}}(i,m), && m\in[1,3],\\
\sum_{(\mu,\mu',m)=\gamma\in\Gamma^{\mathrm{int}}} \mathrm{ec}_{\mathrm{CT}}([\gamma]^{\mathrm{int}}) &= \sum_{k\in[1,k_{\mathrm{C}}]} \delta_{\beta}^{\mathrm{CT}}(k,m), && m\in[1,3],\\
\sum_{(\mu,\mu',m)=\gamma\in\Gamma^{\mathrm{int}}} \mathrm{ec}_{\mathrm{TC}}([\gamma]^{\mathrm{int}}) &= \sum_{k\in[1,k_{\mathrm{C}}]} \delta_{\beta}^{\mathrm{TC}}(k,m), && m\in[1,3],\\
\sum_{(\mu,\mu',m)=\gamma\in\Gamma^{\mathrm{int}}} \mathrm{ec}_{\mathrm{CF}}([\gamma]^{\mathrm{int}}) &= \sum_{c\in[1,\widetilde{t_{\mathrm{C}}}]} \delta_{\beta}^{*\mathrm{F}}(c,m), && m\in[1,3],\\
\sum_{(\mu,\mu',m)=\gamma\in\Gamma^{\mathrm{int}}} \mathrm{ec}_{\mathrm{TF}}([\gamma]^{\mathrm{int}}) &= \sum_{c\in[\widetilde{t_{\mathrm{C}}}+1,c_{\mathrm{F}}]} \delta_{\beta}^{*\mathrm{F}}(c,m), && m\in[1,3],
\end{aligned} \tag{88}$$

$$\begin{aligned}
\sum_{\gamma=(\mathtt{a}d,\mathtt{b}d',m)\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{C}}} [(\mathtt{a},\mathtt{b},m)]^{\mathrm{int}}\cdot\delta_{\mathrm{ec}}^{\mathrm{C}}(i,[\gamma]^{\mathrm{int}}) &= \sum_{\nu\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{C}}} [\nu]^{\mathrm{int}}\cdot\delta_{\mathrm{ac}}^{\mathrm{C}}(i,[\nu]^{\mathrm{int}}),\\
\Delta_{\mathrm{ec}}^{\mathrm{C}+}(i)+\sum_{\gamma=(\mathtt{a}d,\mu',m)\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{C}}} d\cdot\delta_{\mathrm{ec}}^{\mathrm{C}}(i,[\gamma]^{\mathrm{int}}) &= \deg^{\mathrm{C}}(\mathrm{tail}(i)),\\
\Delta_{\mathrm{ec}}^{\mathrm{C}-}(i)+\sum_{\gamma=(\mu,\mathtt{b}d,m)\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{C}}} d\cdot\delta_{\mathrm{ec}}^{\mathrm{C}}(i,[\gamma]^{\mathrm{int}}) &= \deg^{\mathrm{C}}(\mathrm{head}(i)),\\
\Delta_{\mathrm{ec}}^{\mathrm{C}+}(i)+\Delta_{\mathrm{ec}}^{\mathrm{C}-}(i) &\le 8(1-e^{\mathrm{C}}(i)), && i\in[\widetilde{k_{\mathrm{C}}}+1,m_{\mathrm{C}}],\\
\sum_{i\in[\widetilde{k_{\mathrm{C}}}+1,m_{\mathrm{C}}]} \delta_{\mathrm{ec}}^{\mathrm{C}}(i,[\gamma]^{\mathrm{int}}) &= \mathrm{ec}_{\mathrm{C}}([\gamma]^{\mathrm{int}}), && \gamma\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{C}},
\end{aligned} \tag{89}$$

$$\sum_{i\in I_{\mathrm{lnk}}\cap[\widetilde{k_{\mathrm{C}}}+1,m_{\mathrm{C}}]} \delta_{\mathrm{ec}}^{\mathrm{C}}(i,[\gamma]^{\mathrm{int}}) = \mathrm{ec}_{\mathrm{C}}^{\mathrm{lnk}}([\gamma]^{\mathrm{int}}), \qquad \gamma\in\Gamma^{\mathrm{lnk}}\cup\Gamma_{>}^{\mathrm{lnk}}, \tag{90}$$

$$\begin{aligned}
\sum_{\gamma=(\mathtt{a}d,\mathtt{b}d',m)\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{T}}} [(\mathtt{a},\mathtt{b},m)]^{\mathrm{int}}\cdot\delta_{\mathrm{ec}}^{\mathrm{T}}(i,[\gamma]^{\mathrm{int}}) &= \sum_{\nu\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{T}}} [\nu]^{\mathrm{int}}\cdot\delta_{\mathrm{ac}}^{\mathrm{T}}(i,[\nu]^{\mathrm{int}}),\\
\Delta_{\mathrm{ec}}^{\mathrm{T}+}(i)+\sum_{\gamma=(\mathtt{a}d,\mu',m)\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{T}}} d\cdot\delta_{\mathrm{ec}}^{\mathrm{T}}(i,[\gamma]^{\mathrm{int}}) &= \deg^{\mathrm{T}}(i-1),\\
\Delta_{\mathrm{ec}}^{\mathrm{T}-}(i)+\sum_{\gamma=(\mu,\mathtt{b}d,m)\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{T}}} d\cdot\delta_{\mathrm{ec}}^{\mathrm{T}}(i,[\gamma]^{\mathrm{int}}) &= \deg^{\mathrm{T}}(i),\\
\Delta_{\mathrm{ec}}^{\mathrm{T}+}(i)+\Delta_{\mathrm{ec}}^{\mathrm{T}-}(i) &\le 8(1-e^{\mathrm{T}}(i)), && i\in[2,t_{\mathrm{T}}],\\
\sum_{i\in[2,t_{\mathrm{T}}]} \delta_{\mathrm{ec}}^{\mathrm{T}}(i,[\gamma]^{\mathrm{int}}) &= \mathrm{ec}_{\mathrm{T}}([\gamma]^{\mathrm{int}}), && \gamma\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{T}},
\end{aligned} \tag{91}$$

$$
\begin{aligned}
\delta_{\rm ec}^{\rm T}(i,[\gamma]^{\rm int})+\sum_{k\in I_{\rm lnk}\cap[1,k_{\rm C}]}\chi^{\rm T}(i,k) \geq 2\delta_{\rm ec}^{\rm T,lnk}(i,[\gamma]^{\rm int}), \quad & i\in[2,t_{\rm T}],\\
\delta_{\rm ec}^{\rm T,lnk}(i,[\gamma]^{\rm int}) \geq \delta_{\rm ec}^{\rm T}(i,[\gamma]^{\rm int})+\sum_{k\in I_{\rm lnk}\cap[1,k_{\rm C}]}\chi^{\rm T}(i,k)-1, \quad & i\in[2,t_{\rm T}],\\
\sum_{i\in[2,t_{\rm T}]}\delta_{\rm ec}^{\rm T,lnk}(i,[\gamma]^{\rm int}) = {\rm ec}_{\rm T}^{\rm lnk}([\gamma]^{\rm int}), \qquad & \gamma\in\Gamma^{\rm lnk}\cup\Gamma_{>}^{\rm lnk},
\end{aligned}
\tag{92}
$$

$$
\begin{aligned}
\sum_{\gamma=({\tt a}d,{\tt b}d',m)\in\widetilde{\Gamma}_{\rm ec}^{\rm F}}[({\tt a},{\tt b},m)]^{\rm int}\cdot\delta_{\rm ec}^{\rm F}(i,[\gamma]^{\rm int}) = \sum_{\nu\in\widetilde{\Gamma}_{\rm ac}^{\rm F}}[\nu]^{\rm int}\cdot\delta_{\rm ac}^{\rm F}(i,[\nu]^{\rm int}), & \\
\Delta_{\rm ec}^{\rm F+}(i)+\sum_{\gamma=({\tt a}d,\mu',m)\in\widetilde{\Gamma}_{\rm ec}^{\rm F}} d\cdot\delta_{\rm ec}^{\rm F}(i,[\gamma]^{\rm int}) = \deg^{\rm F}(i-1), & \\
\Delta_{\rm ec}^{\rm F-}(i)+\sum_{\gamma=(\mu,{\tt b}d,m)\in\widetilde{\Gamma}_{\rm ec}^{\rm F}} d\cdot\delta_{\rm ec}^{\rm F}(i,[\gamma]^{\rm int}) = \deg^{\rm F}(i,0), & \\
\Delta_{\rm ec}^{\rm F+}(i)+\Delta_{\rm ec}^{\rm F-}(i) \leq 8(1-e^{\rm F}(i)), \qquad & i\in[2,t_{\rm F}],\\
\sum_{i\in[2,t_{\rm F}]}\delta_{\rm ec}^{\rm F}(i,[\gamma]^{\rm int}) = {\rm ec}_{\rm F}([\gamma]^{\rm int}), \qquad & \gamma\in\widetilde{\Gamma}_{\rm ec}^{\rm F},
\end{aligned}
\tag{93}
$$

$$
\begin{aligned}
\deg^{\rm T}(i)+4(1-\chi^{\rm T}(i,k)+e^{\rm T}(i)) \geq \deg_{\rm T}^{\rm CT}(k), & \\
\deg_{\rm T}^{\rm CT}(k) \geq \deg^{\rm T}(i)-4(1-\chi^{\rm T}(i,k)+e^{\rm T}(i)), \qquad & i\in[1,t_{\rm T}],\\
\sum_{\gamma=({\tt a}d,{\tt b}d',m)\in\widetilde{\Gamma}_{\rm ec}^{\rm CT}}[({\tt a},{\tt b},m)]^{\rm int}\cdot\delta_{\rm ec,C}^{\rm CT}(k,[\gamma]^{\rm int}) = \sum_{\nu\in\widetilde{\Gamma}_{\rm ac}^{\rm CT}}[\nu]^{\rm int}\cdot\delta_{\rm ac}^{\rm CT}(k,[\nu]^{\rm int}), & \\
\Delta_{\rm ec}^{\rm CT+}(k)+\sum_{\gamma=({\tt a}d,\mu',m)\in\widetilde{\Gamma}_{\rm ec}^{\rm CT}} d\cdot\delta_{\rm ec,C}^{\rm CT}(k,[\gamma]^{\rm int}) = \deg^{\rm C}({\rm tail}(k)), & \\
\Delta_{\rm ec}^{\rm CT-}(k)+\sum_{\gamma=(\mu,{\tt b}d,m)\in\widetilde{\Gamma}_{\rm ec}^{\rm CT}} d\cdot\delta_{\rm ec,C}^{\rm CT}(k,[\gamma]^{\rm int}) = \deg_{\rm T}^{\rm CT}(k), & \\
\Delta_{\rm ec}^{\rm CT+}(k)+\Delta_{\rm ec}^{\rm CT-}(k) \leq 8(1-\delta_{\chi}^{\rm T}(k)), \qquad & k\in[1,k_{\rm C}],\\
\sum_{k\in[1,k_{\rm C}]}\delta_{\rm ec,C}^{\rm CT}(k,[\gamma]^{\rm int}) = {\rm ec}_{\rm CT}([\gamma]^{\rm int}), \qquad & \gamma\in\widetilde{\Gamma}_{\rm ec}^{\rm CT},
\end{aligned}
\tag{94}
$$

$$
\sum_{i\in I_{\rm lnk}\cap[1,k_{\rm C}]}\delta_{\rm ec}^{\rm CT}(i,[\gamma]^{\rm int}) = {\rm ec}_{\rm CT}^{\rm lnk}([\gamma]^{\rm int}), \qquad \gamma\in\Gamma^{\rm lnk}\cup\Gamma_{>}^{\rm lnk},
\tag{95}
$$

$$
\begin{aligned}
&\deg^{\mathrm{T}}(i) + 4(1 - \chi^{\mathrm{T}}(i,k) + e^{\mathrm{T}}(i+1)) \geq \deg_{\mathrm{T}}^{\mathrm{TC}}(k), \\
&\deg_{\mathrm{T}}^{\mathrm{TC}}(k) \geq \deg^{\mathrm{T}}(i) - 4(1 - \chi^{\mathrm{T}}(i,k) + e^{\mathrm{T}}(i+1)), && i \in [1, t_{\mathrm{T}}], \\
&\sum_{\gamma=(\mathtt{a}d,\mathtt{b}d',m)\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{TC}}} [(\mathtt{a},\mathtt{b},m)]^{\mathrm{int}} \cdot \delta_{\mathrm{ec,C}}^{\mathrm{TC}}(k,[\gamma]^{\mathrm{int}}) = \sum_{\nu\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{TC}}} [\nu]^{\mathrm{int}} \cdot \delta_{\mathrm{ac}}^{\mathrm{TC}}(k,[\nu]^{\mathrm{int}}), \\
&\Delta_{\mathrm{ec}}^{\mathrm{TC+}}(k) + \sum_{\gamma=(\mathtt{a}d,\mu',m)\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{TC}}} d \cdot \delta_{\mathrm{ec,C}}^{\mathrm{TC}}(k,[\gamma]^{\mathrm{int}}) = \deg_{\mathrm{T}}^{\mathrm{TC}}(k), \\
&\Delta_{\mathrm{ec}}^{\mathrm{TC-}}(k) + \sum_{\gamma=(\mu,\mathtt{b}d,m)\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{TC}}} d \cdot \delta_{\mathrm{ec,C}}^{\mathrm{TC}}(k,[\gamma]^{\mathrm{int}}) = \deg^{\mathrm{C}}(\mathrm{head}(k)), \\
&\Delta_{\mathrm{ec}}^{\mathrm{TC+}}(k) + \Delta_{\mathrm{ec}}^{\mathrm{TC-}}(k) \leq 8(1 - \delta_{\chi}^{\mathrm{T}}(k)), && k \in [1, k_{\mathrm{C}}], \\
&\sum_{k\in[1,k_{\mathrm{C}}]} \delta_{\mathrm{ec,C}}^{\mathrm{TC}}(k,[\gamma]^{\mathrm{int}}) = \mathrm{ec}_{\mathrm{TC}}([\gamma]^{\mathrm{int}}), && \gamma \in \widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{TC}},
\end{aligned} \tag{96}
$$

$$
\sum_{i\in I_{\mathrm{lnk}}\cap[1,k_{\mathrm{C}}]} \delta_{\mathrm{ec}}^{\mathrm{TC}}(i,[\gamma]^{\mathrm{int}}) = \mathrm{ec}_{\mathrm{TC}}^{\mathrm{lnk}}([\gamma]^{\mathrm{int}}), \qquad \gamma \in \Gamma^{\mathrm{lnk}} \cup \Gamma_{>}^{\mathrm{lnk}}, \tag{97}
$$

$$
\begin{aligned}
&\deg^{\mathrm{F}}(i) + 4(1 - \chi^{\mathrm{F}}(i,c) + e^{\mathrm{F}}(i)) \geq \deg_{\mathrm{F}}^{\mathrm{CF}}(c), \\
&\deg_{\mathrm{F}}^{\mathrm{CF}}(c) \geq \deg^{\mathrm{F}}(i) - 4(1 - \chi^{\mathrm{F}}(i,c) + e^{\mathrm{F}}(i)), && i \in [1, t_{\mathrm{F}}], \\
&\sum_{\gamma=(\mathtt{a}d,\mathtt{b}d',m)\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{CF}}} [(\mathtt{a},\mathtt{b},m)]^{\mathrm{int}} \cdot \delta_{\mathrm{ec,C}}^{\mathrm{CF}}(c,[\gamma]^{\mathrm{int}}) = \sum_{\nu\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{CF}}} [\nu]^{\mathrm{int}} \cdot \delta_{\mathrm{ac}}^{\mathrm{CF}}(c,[\nu]^{\mathrm{int}}), \\
&\Delta_{\mathrm{ec}}^{\mathrm{CF+}}(c) + \sum_{\gamma=(\mathtt{a}d,\mu',m)\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{CF}}} d \cdot \delta_{\mathrm{ec,C}}^{\mathrm{CF}}(c,[\gamma]^{\mathrm{int}}) = \deg^{\mathrm{C}}(c), \\
&\Delta_{\mathrm{ec}}^{\mathrm{CF-}}(c) + \sum_{\gamma=(\mu,\mathtt{b}d,m)\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{CF}}} d \cdot \delta_{\mathrm{ec,C}}^{\mathrm{CF}}(c,[\gamma]^{\mathrm{int}}) = \deg_{\mathrm{F}}^{\mathrm{CF}}(c), \\
&\Delta_{\mathrm{ec}}^{\mathrm{CF+}}(c) + \Delta_{\mathrm{ec}}^{\mathrm{CF-}}(c) \leq 8(1 - \delta_{\chi}^{\mathrm{F}}(c)), && c \in [1, \widetilde{t_{\mathrm{C}}}], \\
&\sum_{c\in[1,\widetilde{t_{\mathrm{C}}}]} \delta_{\mathrm{ec,C}}^{\mathrm{CF}}(c,[\gamma]^{\mathrm{int}}) = \mathrm{ec}_{\mathrm{CF}}([\gamma]^{\mathrm{int}}), && \gamma \in \widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{CF}},
\end{aligned} \tag{98}
$$

$$
\begin{aligned}
&\deg^{\mathrm{F}}(j) + 4(1 - \chi^{\mathrm{F}}(j,i+\widetilde{t_{\mathrm{C}}}) + e^{\mathrm{F}}(j)) \geq \deg_{\mathrm{F}}^{\mathrm{TF}}(i), \\
&\deg_{\mathrm{F}}^{\mathrm{TF}}(i) \geq \deg^{\mathrm{F}}(j) - 4(1 - \chi^{\mathrm{F}}(j,i+\widetilde{t_{\mathrm{C}}}) + e^{\mathrm{F}}(j)), && j \in [1, t_{\mathrm{F}}], \\
&\sum_{\gamma=(\mathtt{a}d,\mathtt{b}d',m)\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{TF}}} [(\mathtt{a},\mathtt{b},m)]^{\mathrm{int}} \cdot \delta_{\mathrm{ec,T}}^{\mathrm{TF}}(i,[\gamma]^{\mathrm{int}}) = \sum_{\nu\in\widetilde{\Gamma}_{\mathrm{ac}}^{\mathrm{TF}}} [\nu]^{\mathrm{int}} \cdot \delta_{\mathrm{ac}}^{\mathrm{TF}}(i,[\nu]^{\mathrm{int}}), \\
&\Delta_{\mathrm{ec}}^{\mathrm{TF+}}(i) + \sum_{\gamma=(\mathtt{a}d,\mu',m)\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{TF}}} d \cdot \delta_{\mathrm{ec,T}}^{\mathrm{TF}}(i,[\gamma]^{\mathrm{int}}) = \deg^{\mathrm{T}}(i), \\
&\Delta_{\mathrm{ec}}^{\mathrm{TF-}}(i) + \sum_{\gamma=(\mu,\mathtt{b}d,m)\in\widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{TF}}} d \cdot \delta_{\mathrm{ec,T}}^{\mathrm{TF}}(i,[\gamma]^{\mathrm{int}}) = \deg_{\mathrm{F}}^{\mathrm{TF}}(i), \\
&\Delta_{\mathrm{ec}}^{\mathrm{TF+}}(i) + \Delta_{\mathrm{ec}}^{\mathrm{TF-}}(i) \leq 8(1 - \delta_{\chi}^{\mathrm{F}}(i+\widetilde{t_{\mathrm{C}}})), && i \in [1, t_{\mathrm{T}}], \\
&\sum_{i\in[1,t_{\mathrm{T}}]} \delta_{\mathrm{ec,T}}^{\mathrm{TF}}(i,[\gamma]^{\mathrm{int}}) = \mathrm{ec}_{\mathrm{TF}}([\gamma]^{\mathrm{int}}), && \gamma \in \widetilde{\Gamma}_{\mathrm{ec}}^{\mathrm{TF}},
\end{aligned} \tag{99}
$$

$$\begin{aligned}\sum_{\mathrm{X}\in\{\mathrm{C,T,F,CT,TC,CF,TF}\}}(\mathrm{ec}_{\mathrm{X}}([\gamma]^{\mathrm{int}})+\mathrm{ec}_{\mathrm{X}}([\overline{\gamma}]^{\mathrm{int}}))=\mathrm{ec}^{\mathrm{int}}([\gamma]^{\mathrm{int}}), &\quad \gamma\in\Gamma^{\mathrm{int}}_{<},\\ \sum_{\mathrm{X}\in\{\mathrm{C,T,F,CT,TC,CF,TF}\}}\mathrm{ec}_{\mathrm{X}}([\gamma]^{\mathrm{int}})=\mathrm{ec}^{\mathrm{int}}([\gamma]^{\mathrm{int}}), &\quad \gamma\in\Gamma^{\mathrm{int}}_{=},\end{aligned} \tag{100}$$

$$\begin{aligned}\sum_{\mathrm{X}\in\{\mathrm{C,T,CT,TC}\}}(\mathrm{ec}^{\mathrm{lnk}}_{\mathrm{X}}([\gamma]^{\mathrm{int}})+\mathrm{ec}^{\mathrm{lnk}}_{\mathrm{X}}([\overline{\gamma}]^{\mathrm{int}}))=\mathrm{ec}^{\mathrm{lnk}}([\gamma]^{\mathrm{int}}), &\quad \gamma\in\Gamma^{\mathrm{lnk}}_{<},\\ \sum_{\mathrm{X}\in\{\mathrm{C,T,CT,TC}\}}\mathrm{ec}^{\mathrm{lnk}}_{\mathrm{X}}([\gamma]^{\mathrm{int}})=\mathrm{ec}^{\mathrm{lnk}}([\gamma]^{\mathrm{int}}), &\quad \gamma\in\Gamma^{\mathrm{lnk}}_{=}.\end{aligned} \tag{101}$$

$$\sum_{\gamma\in\Gamma^{\mathrm{lnk}}}\mathrm{ec}^{\mathrm{lnk}}([\gamma]^{\mathrm{int}})=n_{\mathrm{lnk}}, \tag{102}$$

$$\mathrm{ns}^{\mathrm{cnt}}_{\mathrm{LB}}([\mu])\le\delta^{\mathrm{cnt}}(1,[\mu])+\delta^{\mathrm{cnt}}(2,[\mu])\le\mathrm{ns}^{\mathrm{cnt}}_{\mathrm{UB}}([\mu]), \quad \mu\in\Lambda^{\mathrm{int}}_{\mathrm{dg}}, \tag{103}$$

$$\begin{aligned}\sum_{\gamma\in\Gamma^{\mathrm{cnt}}_{<}\cup\Gamma^{\mathrm{cnt}}_{=}\cup\Gamma^{\mathrm{cnt}}_{>}}\delta^{\mathrm{cnt}}([\gamma]^{\mathrm{int}})=1, &\\ \mathrm{ec}^{\mathrm{lnk}}([\gamma]^{\mathrm{int}})\ge\delta^{\mathrm{cnt}}([\gamma]^{\mathrm{int}}), &\quad \gamma\in\Gamma^{\mathrm{cnt}}_{<}\cup\Gamma^{\mathrm{cnt}}_{=}\\ \mathrm{ec}^{\mathrm{lnk}}([\overline{\gamma}]^{\mathrm{int}})\ge\delta^{\mathrm{cnt}}([\gamma]^{\mathrm{int}}), &\quad \gamma\in\Gamma^{\mathrm{cnt}}_{>}\end{aligned} \tag{104}$$

5.11 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 $x = (x(1), x(2), \ldots, x(K))$:

$$\frac{(1-\varepsilon)(x(j)-\min(\mathrm{dcp}_j;D_\pi))}{\max(\mathrm{dcp}_j;D_\pi)-\min(\mathrm{dcp}_j;D_\pi)}\le\widehat{x}(j)\le\frac{(1+\varepsilon)(x(j)-\min(\mathrm{dcp}_j;D_\pi))}{\max(\mathrm{dcp}_j;D_\pi)-\min(\mathrm{dcp}_j;D_\pi)},\ j\in[1,K]. \tag{105}$$

An example of a tolerance is $\varepsilon = 1 \times 10^{-5}$.

We use the same conversion for descriptor $x_j = \overline{\mathrm{ms}}$.

References

[1] Zhu, J., Azam, N. A., Haraguchi, K., Zhao, L., Nagamochi, H., Akutsu, T.: A method for molecular design based on linear regression and integer programming, arXiv: 2107.02381 (2021).

[2] Hoerl, A., Kennard, R.: Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, **12**(1), 55–67 (1970)

[3] Hoerl, A., Kennard, R.: Ridge Regression: Applications to Nonorthogonal Problems. Technometrics, **12**(1), 69–82 (1970)

[4] Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Statist. Soc. B **58**, 267–288 (1996)