Topology-Aware Hierarchical Graph Diffusion Model for Molecular Graph Generation

Rongshen He[1], Abubakar Zakari[2], Qinru Yang[1], Jiaqi Luo[1], Changsheng Ma[1](✉)

[1] School of Information Science and Engineering, Lanzhou University
{320220940271, 320220940961, 320220917431, macs}@lzu.edu.cn
[2] Fashable AI, Quinta da Barca, R. Nossa Sra. de Guadalupe 113 Lote 78, Loja 1, 4740-473 Esposende, Portugal {abubakar}@fashable.ai

Abstract. This work introduces THGD, a Topology-Aware Hierarchical Graph Diffusion Model designed to address the challenges of generating large, structurally complex molecules. THGD employs a coarse-to-fine framework that decouples global topology preservation from local atomic refinement, enabling precise generative control and efficient exploration of broader chemical spaces without relying on restrictive, predefined motif vocabularies. Extensive experiments underscore THGD's superior performance. It robustly preserves complex structural constraints, achieving up to 2× higher scaffold validity than the previous state-of-the-art model in scaffold-constrained generation task. Furthermore, in molecular generation task, THGD excels in generating large molecules with high distribution fidelity, attaining an FCD score of 80.26 on the challenging GuacaMol dataset, effectively matching the diversity of real-world molecular distributions. These results highlight THGD's potential to advance molecular design for drug discovery and beyond. Our code is available at https://github.com/hers22/THGD.

Keywords: Molecular Generation · Hierarchical Diffusion · Graph Coarsening · Large Molecules

1 Introduction

Modern drug discovery requires the development of advanced machine learning models capable of effectively capturing and sampling from the vast chemical space. This task is both highly challenging and urgent, given the astronomical scale of drug-like compounds, estimated to exceed 10^{60} [18]. Recently, deep graph generative models, which model the underlying probability distributions of molecular graph structures, enabling the generation of high-quality candidate molecules while avoiding the prohibitive costs associated with exhaustive chemical space searches [6, 27], have achieved significant success in this domain, demonstrating their potential to revolutionize molecular design.

Although existing graph generative models have shown promising results, significant challenges remain in generating large and structurally complex molecules.

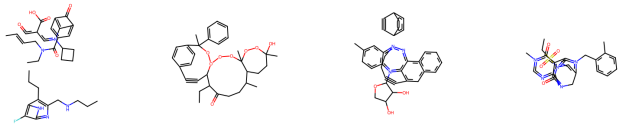

Fig. 1: Visualization results of unreasonable structure generated by atom-based generative methods.

These models can be broadly categorized into two main paradigms: atom-based generative models [21] and motif-based generative models [9]. Atom-based generative models construct molecules at the atom level, treating each atom as a node and each bond as an edge in a graph. While this fine-grained approach allows for precise control over molecular structures, it suffers from scalability issues. As the size of the molecular graph grows, the computational complexity increases drastically due to the quadratic scaling of edges with respect to nodes. This leads to substantial computational overhead, escalating memory requirements, and difficulties in accurately predicting sparse chemical bonds, often resulting in disconnected substructures, implausible ring systems, and invalid valency states, as illustrated in Figure 1.

On the other hand, motif-based generative models construct molecules by assembling predefined structural motifs, such as functional groups or subgraphs. By leveraging larger building blocks, this approach reduces combinatorial complexity and improves scalability compared to atom-based models. However, these models are constrained by their heavy reliance on predefined motif libraries, which inherently limits the exploration of the chemical space and the ability to generate truly innovative molecules that deviate from predefined patterns. While enlarging the motif library might alleviate this issue, it complicates the modeling of diverse motif graphs due to their permutation-unequivariant nature [24].

To address these challenges, we introduce THGD, a novel topology-aware hierarchical graph diffusion model designed for large molecular graph generation. THGD employs a coarse-to-fine hierarchical diffusion process that decouples global topology preservation from local atomic refinement, enabling precise control over molecular generation while maintaining scalability. The model first generates a coarse graph representation through spectral-preserved graph coarsening, dynamically identifying and partitioning ring structures without relying on predefined motif libraries.

The coarse graph is then refined into detailed atom level structures using a conditioned diffusion process guided by structural type priors, ensuring chemically realistic substructure generation. Extensive experiments conducted on the GuacaMol [2] and MOSES [20] benchmarks demonstrate THGD's superior performance in capturing true dataset distributions and excelling in scaffold-constrained generation tasks. The contributions of this work are as follows:

1. **Novel Hierarchical Framework:** We propose a coarse-to-fine diffusion model that decouples global topology preservation from local atomic refine-

ment, addressing scalability and chemical validity challenges faced by existing methods.
2. **Elimination of Predefined Motif Libraries:** By leveraging spectral-preserved graph coarsening and type-specific marginal priors, THGD eliminates reliance on predefined motif libraries, enhancing flexibility and generalization.
3. **State-of-the-Art Performance:** THGD achieves state-of-the-art results on both generation and scaffold-constrained generation tasks, with an FCD score of 80.26 on GuacaMol and up to 2× higher scaffold validity compared to previous state-of-the-art model.

2 Related Work

Molecule generation remains a cornerstone challenge in drug discovery. Early approaches relied on SMILES-based sequence generation models [3, 5, 12], which pioneered automated molecular design but struggled to explicitly model chemical topology and valency rules, resulting in limited validity rates for complex molecules. Subsequent research shifted to molecular graph representations, achieving significant progress through GANs [4], VAEs [15], and normalizing flows [27].

Recent breakthroughs in diffusion and score-based models have revolutionized the field. For instance, EDM [7] introduced permutation-equivariant diffusion for 3D molecular conformations, GDSS [11] leveraged stochastic differential equations to model node and edge features, and DiGress [24] advanced discrete diffusion techniques. Building on these foundations, Cometh [22] and DisCo [25] reformulated discrete diffusion modeling using continuous-time Markov chains [1]. While existing methods achieve strong performance in generating small molecules, they exhibit significant limitations in fitting the distribution of large molecules due to combinatorial explosion and strict chemical valence rules, particularly as molecular complexity escalates exponentially.

To address scalability, hierarchical frameworks decompose molecular generation into coarse-to-fine stages. Existing methods such as tree decompositions [9] and atom-motif hierarchies [10] reduce complexity by leveraging predefined structural motifs. However, their reliance on fixed motif libraries restricts the exploration of novel chemical spaces. This limitation is commonly observed in structured data generation tasks, where static prior knowledge hinders adaptability to dynamic patterns [26]. In contrast, our proposed method employs a hierarchical framework that eliminates dependence on predefined vocabularies through spectral-preserved graph coarsening and dynamic structural typing, enabling broader exploration of the chemical space while maintaining scalability and validity.

3 Background

Graph Diffusion Models Diffusion models [6, 23] are probabilistic generative models defined by a forward diffusion process that gradually adds noise to

graphs and a reverse denoising process that learns to remove it, enabling sample generation. Formally, the model is expressed as a latent variable model:

$$p_\theta(G_0) := \int p_\theta(G_{0:T})\, dG_{1:T}. \tag{1}$$

Recent advancements extend denoising diffusion probabilistic models to discrete domains, with DiGress [24] applying this framework to graph G with its nodes V and edges E. Noise is added to graphs using transition matrices Q^t, where $[Q^t]_{ij}$ defines transitions from state i to j. The noisy graph G^t in the forward process is computed as:

$$q(G^t|G_0) = G_0\bar{Q}^t, \quad \text{with} \quad \bar{Q}^t = \bar{\alpha}_t I + (1-\bar{\alpha}_t)\mathbf{1}m^\top, \tag{2}$$

where $\bar{\alpha}_t$ controls noise levels, and m approximates the true data distribution $q_V \times q_E$. Transition matrices satisfy:

$$\lim_{T\to\infty} \bar{Q}_T^\top \mathbf{1} = m, \tag{3}$$

ensuring transitions align with training set marginal probabilities.

The reverse process reconstructs the clean graph iteratively using Bayes' theorem:

$$q(G^{t-1}|G^t, G_0) \propto G^t (Q^t)^\top \odot G_0 \bar{Q}^{t-1}, \tag{4}$$

where $(Q^t)^\top$ is the transpose of Q^t, and $\odot$ denotes element-wise multiplication.

A denoising neural network ϕ_θ predicts clean graphs from noisy inputs $G^t = (V^t, E^t)$. It is trained by minimizing the cross-entropy loss between predicted probabilities $\hat{p}^G = (\hat{p}^V, \hat{p}^E)$ and ground truth G:

$$l(\hat{p}^G, G) = \sum_i \mathrm{CE}(v_i, \hat{p}_i^V) + \lambda \sum_{i,j} \mathrm{CE}(e_{ij}, \hat{p}_{ij}^E), \tag{5}$$

where $\lambda \in \mathbb{R}^+$ balances node and edge importance, and CE denotes cross-entropy.

4 Methodology

We propose THGD, a novel hierarchical molecular generation framework that employs a two-stage diffusion process to synthesize molecules at both coarse topological and fine atomic levels. As illustrated in Figure 2, the first stage trains a discrete diffusion model to generate coarse graphs representing cluster level molecular topologies. The second stage refines these coarse graphs into detailed atom level structures using a conditioned diffusion model. This decomposition enables efficient generation while preserving critical chemical constraints.

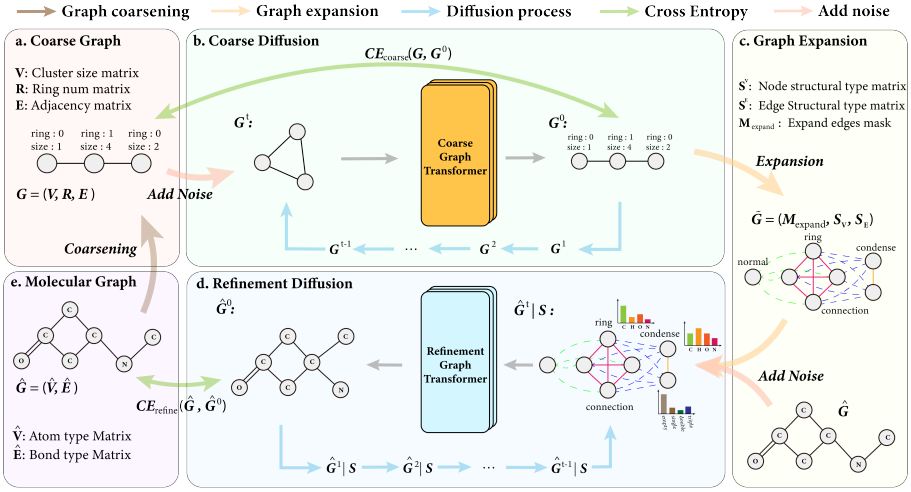

Fig. 2: Architecture of *THGD*: (I) Coarse graph generation via discrete diffusion learns cluster-level topologies; (II) Refined graph generation uses topology-conditioned diffusion to reconstruct atom-level details. Both stages employ graph transformers trained with cross-entropy loss.

4.1 Notation and Definitions

A molecular graph $\hat{G} = (\hat{\boldsymbol{V}}, \hat{\mathbf{E}})$ containing $\hat{n}$ atoms also noted as refined graph, represents the original molecular structure. The node features $\hat{\boldsymbol{V}} \in \{0,1\}^{\hat{n}\times a}$ are one-hot encodings of atom types, where a is the total number of distinct atom types. Similarly, edge features $\hat{\boldsymbol{E}} \in \{0,1\}^{\hat{n}\times\hat{n}\times b}$ are one-hot encodings of bond types between pairs of atoms, where b is the total number of distinct bond types.

A coarse graph $G = (\boldsymbol{V}, \boldsymbol{R}, \boldsymbol{E})$ is an abstracted representation of the molecular topology, consisting of n clusters (where $n \leq \hat{n}$). The cluster node features $\boldsymbol{V} \in \{0,1\}^{n\times c}$ are one-hot encodings representing the number of atoms in each cluster, where c is the maximum number of atoms a single cluster can contain. The cluster ring features $\boldsymbol{R} \in \{0,1\}^{n\times r}$ are one-hot encodings that represents the number of fused rings within each cluster, where r is the maximum number of fused rings a single cluster can represent. The adjacency matrix $\boldsymbol{E} \in \{0,1\}^{n\times n}$ encodes inter-cluster connectivity. This abstraction simplifies the molecular graph while preserving its high-level topological structure.

An expanded graph $\tilde{G} = (\boldsymbol{M}_{\text{expand}}, \boldsymbol{S}_{\boldsymbol{V}}, \boldsymbol{S}_{\boldsymbol{E}})$ serves as an intermediate representation during the refinement process, bridging the coarse graph G and the molecular graph $\hat{G}$. It has $\tilde{n}$ nodes, where $\tilde{n} = \hat{n}$ to align with the atom count of the original molecular graph. It employs a connectivity mask $\boldsymbol{M}_{\text{expand}} \in \{0,1\}^{\tilde{n}\times\tilde{n}}$, which encodes permissible atom-atom connections derived from the expanded clusters, enforcing topological constraints from the coarse graph. Furthermore, it integrates structural matrices $\boldsymbol{S}_{\boldsymbol{V}} \in \{0,1\}^{\tilde{n}\times s}$ and $\boldsymbol{S}_{\boldsymbol{E}} \in \{0,1\}^{\tilde{n}\times\tilde{n}\times s}$, which

are hot encodings that represent the structural types of the nodes and edges derived from the coarse graph, where s denotes the possible structural types and is consistent for both nodes and edges.

4.2 Coarsening and Refinement

Traditional motif-based coarsening methods rely on predefined motif vocabularies, which constrain the exploration of chemical space and impede the generation of innovative molecules. To overcome this limitation, we introduce a novel molecular graph coarsening method that eliminates dependence on fixed motif libraries.

Coarsening: The coarsening stage transforms the original molecular graph $\hat{G}$ into a compact coarse graph G, reducing complexity while retaining key chemical and topological information. The quality of the coarse graph is pivotal, we employ a two-stage strategy (illustrated in Fig. 3):

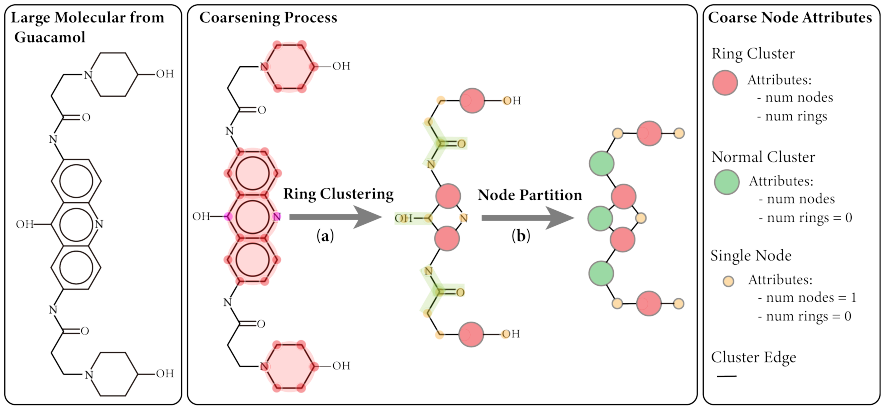

Fig. 3: **Illustration of structure preserved coarsening.** A large molecule from Guacamol dataset was taken as an example. (a) All nodes in rings are clustered together (highlighted in red), and connected rings were partitioned by splitting rings (split nodes in purple); (b) Remaining nodes (highlighted in yellow) are partitioned by spectrum reduction.

1. **Ring Identification and Partitioning:** We begin by identifying the rings in the molecular graph, as they contribute significantly to the structure and properties of the molecule. Fundamental ring structures are identified and merged into clusters. Small rings with fewer than 6 atoms or those sharing atoms with neighboring rings are grouped. To manage complexity from very large ring systems (e.g., up to 14 fused rings, details in Appendix A), we partition them by splitting shared atoms, ensuring manageable cluster sizes. (see Fig. 3, "(a)Ring Clustering").

2. **Graph Coarsening:** The remaining non-ring nodes are partitioned into clusters using spectral-preserved techniques [14], leveraging the Laplacian spectrum to maintain the original graph's topology (Fig. 3b). This ensures the coarse graph accurately reflects the molecular structure, and this dynamic, vocabulary-free approach allows for flexible abstraction of diverse chemical entities. (see Fig. 3, "(b)Node Partition").

Refinement: The refinement stage focuses on reconstructing the original molecular graph from the coarse graph G. However, the coarse graph (with n clusters) and the molecular graph $\hat{G}$ (with $\hat{n}$ atoms) exhibit size misalignment, necessitating joint modeling of distributions across different scales. To address this, we introduce an intermediate expanded graph $\tilde{G} = (M_{\text{expand}}, S_V, S_E)$ as a bridge between the abstracted coarse topology and fine-grained molecular details, ensuring size consistency and preserving chemical information via three interconnected mechanisms: (1) **cluster expansion** to align node counts, (2) **edge masking** (M_{expand}) to regulate connectivity, and (3) **structural typing** (S_V, S_E) to enforce chemical constraints.

1. **Cluster Expansion:** The process begins by expanding each cluster $c_p \in G$ into a set of atomic nodes. Specifically, a cluster containing $V[p]$ atoms is mapped to $V[p]$ atomic nodes through the operator $\text{Expand}(c_p)$, defined as:

$$V' = \bigcup_{p=1}^{n} \text{Expand}(c_p), \quad \text{where } |\text{Expand}(c_p)| = V[p] \text{ and } \sum_{p=1}^{n} V[p] = \tilde{n}.$$

   Here, $\text{Expand}(c_p)$ represents the operator mapping cluster c_p to $V[p]$ atomic nodes, and V' is the expanded node set with size $\tilde{n} = \hat{n}$, matching the molecular graph. This expansion ensures $|V'| = \hat{n}$, aligning $\tilde{G}$ with the molecular graph size while preserving cluster-level information.
2. **Edge Masking:** To regulate connectivity during refinement, we define a binary mask $M_{\text{expand}} \in \{0,1\}^{\tilde{n}\times\tilde{n}}$. This mask encodes permissible edges as:

$$M_{\text{expand}}^{(i,j)} = \begin{cases} 1, & \text{if } v_i, v_j \text{ belong to the same cluster (intra-cluster)} \\ 1, & \text{if } v_i \in c_p, v_j \in c_q \text{ and } E_{p,q} = 1 \text{ (inter-cluster)} \\ 0, & \text{otherwise.} \end{cases}$$

   The mask acts as a structural scaffold: intra-cluster edges preserve local substructures (e.g., aromatic rings), inter-cluster edges enforce connectivity defined in G, and masked regions ($M_{\text{expand}}^{(i,j)} = 0$) prohibit chemically invalid bonds during refinement.
3. **Structural Typing:** To enforce chemical validity, we introduce structural type matrices S_V (nodes) and S_E (edges). The node structural matrix $S_V \in \mathbb{N}^{\tilde{n}\times s}$ assigns each expanded node a type inherited from its parent cluster as:

$$S_V[i] = \begin{cases} \text{NORMAL}, & \text{if } V[p] = 1 \quad \text{(singleton cluster)} \\ \text{CONDENSED}, & \text{if } V[p] > 1 \text{ and } R[p] = 0 \quad \text{(chain/functional group)} \\ \text{RING}_k, & \text{if } R[p] = k \geq 1 \quad \text{(fused ring system).} \end{cases}$$

And the edge structural matrix $S_E \in \mathbb{N}^{\tilde{n}\times\tilde{n}\times s}$ labels edges based on connectivity as:

$$S_E[i,j] = \begin{cases} \text{CONNECTION}, & \text{if } M^{(i,j)}_{\text{expand}} = 1 \text{ and inter-cluster} \\ \text{Inherited from } S_V[i] \text{ or } S_V[j], & \text{if intra-cluster.} \end{cases}$$

Here, s is the number of structural types (NORMAL, CONDENSED, $RING_k$, CONNECTION). These matrices enforce domain-specific constraints. For example, $RING_k$ clusters enforce aromaticity, CONDENSED clusters bias atom types toward chain-appropriate elements, and CONNECTION edges restrict bonds to single/double/triple types.

The expanded graph $\tilde{G}$ guides molecular generation through a conditioned diffusion model that iteratively denoises the noisy molecular graph. This process is guided by structural type priors (S_V, S_E) and the connectivity mask (M_{expand}) derived from the coarse graph. Specifically, S_V/S_E biases atom/bond predictions, while masked regions ($M_{\text{expand}} = 0$) prohibit invalid bonds and enforce valid atom valency during denoising. By decoupling topology preservation from atomic refinement, this hierarchical design ensures alignment with both global molecular topology and local chemical rules, enabling scalable synthesis of complex molecules with chemical validity, as shown in Fig. 4.

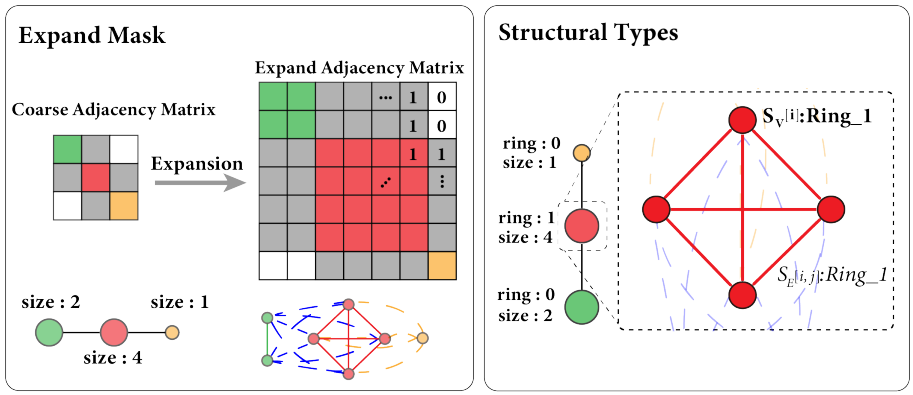

Fig. 4: Illustration of Coarse Graph Expansion.

4.3 Hierarchical Diffusion

Our hierarchical generation process decomposes molecular synthesis into two coupled diffusion stages governed by the joint distribution:

$$q(\hat{G}, G) = \underbrace{q(\hat{G}|G)}_{\text{Atomic Refinement}} \underbrace{q(G)}_{\text{Coarse Topology}} \tag{6}$$

We employ two graph diffusion models to learn both the distribution $q(G)$ of the coarse graph and the conditioned distribution $q(\hat{G} \mid G)$ of the molecular graph.

Coarse Graph Diffusion: Our coarse graph diffusion process is conceptually similar to the approach used in DiGress [24]. Specifically, we model the distribution $q(G)$ of the coarse graph $G = (V, R, E)$ through discrete diffusion over its components. This similarity allows us to directly leverage the established framework of DiGress for learning and sampling coarse graphs, ensuring a solid foundation for our hierarchical generation process.

The training objective for the coarse graph diffusion model is formulated as:

$$L_{\text{coarse}} = \sum_{i=1}^{n} \text{CE}(v_i, \hat{p}_V^i) + \zeta \cdot \text{CE}(r_i, \hat{p}_R^i) + \gamma \cdot \sum_{ij} \text{CE}(e_{ij}, \hat{p}_E^{ij}), \tag{7}$$

where: v_i, r_i, and e_{ij} represent the node features, ring counts, and edge connectivity of the coarse graph, respectively. $\hat{p}_i^V$, $\hat{p}_i^R$, and $\hat{p}_{ij}^E$ are the predicted probabilities for the corresponding attributes. $\zeta \in \mathbb{R}^+$ and $\gamma \in \mathbb{R}^+$ are hyperparameters that balance the importance of ring count prediction, node count prediction, and edge connectivity relative to node attributes.

This formulation ensures that the coarse graph diffusion model effectively captures both the structural and topological characteristics of the molecular graph at the cluster level. By building on the success of DiGress, we achieve robust performance in generating coarse graphs that accurately reflect the underlying molecular topology.

Optimal Marginal Prior Probability: The choice of initial distributions significantly impacts model performance, as demonstrated in DiGress [24], where better priors lead to markedly improved results. We observe a similar phenomenon in our framework: atom and bond types exhibit distinct distributions across different topological structures. For instance, aromatic bonds occur more frequently in ring systems compared to chain structures (see Appendix B for detailed statistics). We compute optimal prior distributions tailored to each structural condition to enhance the discrete diffusion process.

Formally, let M denote the optimal prior distribution for molecular graph generation $q(\hat{V}, \hat{E} \mid S)$, constrained by structural typologies that best approximate the true data distribution. The accumulated node attribute transition matrices $\bar{Q}_T^V$ satisfy:

$$\lim_{T \to \infty} \bar{Q}_T^V \mathbf{1}_a = (S_k^V M^V)^\top, \tag{8}$$

where the structural type matrix $S^V \in \mathbb{N}^{\hat{n} \times s}$ contains node-wise structural types with S_k^V specifying the k-th node's structural type ($k \in [0, \hat{n})$). Here $\mathbf{1}_a$ represents an a-dimensional vector of ones, and $M^V \in \mathbb{R}^{s \times a}$ constitutes the core marginal probability matrix where s enumerates structural categories and a denotes node label cardinality. The probability of transitioning from state i to state j for the k-th refined node is proportional to the marginal probability of M^V selected by the structural type at that node (i.e. $S_k^V M^V$), with a similar formulation applied to edges.

This ensures that the forward process begins closer to the true data distribution, maintaining both unbiased sampling and chemical fidelity throughout the generation process. By incorporating these structure-specific priors, we align the diffusion process with realistic molecular configurations, improving both the quality and efficiency of the generated outputs.

Forward Process: We define a conditional diffusion process in which the coarse graph guides noise injection into the molecular graph. The forward process is formulated as

$$q(\hat{G}_1, \ldots, \hat{G}_T \mid G) = \prod_{t=1}^{T} q(\hat{G}_t \mid \hat{G}_{t-1}, G), \tag{9}$$

where the coarse graph G is then expanded to the expanded graph represented by structural type matrix $\boldsymbol{S}$ that contains topological information and constraints. Consequently, the transition probability can be equivalently written as $q(\hat{G}_t \mid \hat{G}_{t-1}, G) = q(\hat{G}_t \mid \hat{G}_{t-1}, \boldsymbol{S})$.

For each node of the molecular graph, the forward transition matrix for the k-th entry ($k \in [0, \hat{n})$) is defined as

$$Q_k^t = (1 - \beta_t)I + \beta_t \mathbf{1}_a (S_k M), \tag{10}$$

where I denotes the identity matrix, $\mathbf{1}_a$ is an a-dimensional vector of ones and β_t is a time-dependent parameter controlling the noise level. The cumulative transition matrix over t time steps is given by $\bar{Q}^t = Q^1 Q^2 \cdots Q^t$. A similar formulation is applied to the edge features.

Reverse Process: The reverse diffusion process reconstructs the original molecular graph $\hat{G}_0$ from its noisy counterpart $\hat{G}_T$. By applying Bayes' rule, the conditional probability for the reverse step can be expressed as:

$$q(\hat{G}_{t-1} \mid \hat{G}_t, S) = q(\hat{G}_{t-1} \mid \hat{G}_t, \hat{G}_0, S) \propto q(\hat{G}_t \mid \hat{G}_{t-1}, S)\, q(\hat{G}_{t-1} \mid \hat{G}_0, S), \tag{11}$$

where S is the structural type matrix. In this framework, the only modification compared to the standard reverse process is the explicit conditioning on S, which is integrated into the forward transition matrix Q and the input feature. As a result, the reverse process formulation remains consistent with Equation 4. The model thus denoises the noisy molecular graph based on the structural types derived from the coarse graph, ensuring the generation of chemically plausible molecules.

Training refinement denoising network: The refinement denoising neural network $\phi_{\theta r}$ parametrized by θr take taking a noisy molecular graph $\hat{G}_t = (\hat{V}, \hat{E})$ at t-th step conditioned on the coarse graph and the structural types matrices (S_V, S_E) as input and aims to predict the clean refined graph $\hat{G}_0 = (\hat{V}, \hat{E})$. It was trained by optimizing the cross-entropy loss l between the predicted probabilities $\hat{p}^{\hat{G}} = (\hat{p}^{\hat{V}}, \hat{p}^{\hat{E}})$ for each node and edge and the true refined graph following [24]. Notably, the refinement is restricted to chemically plausible regions using an expansion mask M_{expand} attained from expanding the coarse graph, the

refinement loss then focuses only on mask-valid edges:

$$\mathcal{L}_{\text{refine}} = \sum_{i \in \hat{V}} \text{CE}(v_i, \hat{p}^{\hat{V}}) + \lambda \sum_{(i,j) \in M_{\text{expand}}} \text{CE}(e_{ij}, \hat{p}^{\hat{E}}), \tag{12}$$

where $\lambda \in \mathbb{R}^+$ is a hyperparameter balancing the importance of node attributes relative to edge attributes.

4.4 Model review

The overall structure of our proposed model is as shown in Figure 2, which uses a two-stage process: coarse graph generation for high-level structure and expansion for detailed atom-level refinement. Two diffusion models drive these stages, enabling fine-grained control and preserving global structure. Detailed architectures are in Appendix C.
Stage I: Coarse graph generation via discrete diffusion learns cluster-level topologies.

1. **Coarsening:** Abstracts the molecular graph into a coarse graph G^0, where nodes are atom clusters and edges are cluster connections.
2. **Coarse Diffusion Model Training:** Progressively adds noise to G^0 over timesteps, resulting in noisy G^t. A graph transformer then learns to reverse the noise, denoising G^t to predict G^0. Training minimizes cross-entropy loss.
3. **Coarse Diffusion Model Sampling:** Sample a noisy representation G^t from the prior distribution, then progressively remove the noise to G^0 using the learned coarse graph transformer.

Stage II: Refined graph generation uses topology-conditioned diffusion to reconstruct atom-level details.

1. **Expansion and Refinement**: Expand coarse graph nodes into atom clusters with structural types and expand mask.
2. **Refinement Diffusion Model Training:** Progressively adds noise to $\hat{G}$ over timesteps, conditioned by structural types retrieved from expanding G^0. A refinement graph transformer then denoises the expanded graph using structural type information to predict the refined molecular graph. Training minimizes cross-entropy loss under the constrain of expand mask.
3. **Refinement Diffusion Model Sampling:** Sample a noisy representation from the prior distribution conditioned by structural types retrieved from expanding G^0. A refinement graph transformer then progressively removes the noise to $\hat{G}^0$.

This hierarchical approach decomposes molecular generation, allowing THGD to generate complex, valid, and diverse molecules using diffusion models at coarse and fine levels.

4.5 Scaffold Constrained Generation

In drug discovery, generating molecules with specific scaffolds—subgraphs possessing desired chemical properties—is essential. Given a subgraph $S = (V_S, E_S)$ with n_s nodes, the generation process conditioned on S can be achieved by masking the first n_s-th node and edge feature tensors at each reverse iteration step using a permutation-equivariant model [24]. After sampling G^{t-1}, the node and edge features are updated as follows:

$$V^{t-1} = M_V \odot V_s + (1 - M_V) \odot V^{t-1}, \quad E^{t-1} = M_E \odot E_s + (1 - M_E) \odot E^{t-1}, \tag{13}$$

where $M_V \in \{0, 1\}^n$ and $M_E \in \{0, 1\}^{n \times n}$ are binary masks that identify the first n_s nodes and their associated edges, ensuring the preservation of the scaffold structure during the generation process. For our case, scaffold masking was applied separately to both the coarse and refined graphs during their respective generation stages.

5 Experiments

This section provides a thorough evaluation of THGD on both molecular graph generation and scaffold-constrained generation.

We evaluate our model on two large-scale molecular datasets: MOSES [20] and GuacaMol [2]. MOSES is a refined subset of the ZINC database, containing 1.9 million molecules, with 1.6 million designated for training, and graph size averaging 21.7 nodes. GuacaMol is a large molecular dataset derived from the ChEMBL database with an average molecular graph size of 27.8 nodes, includes 1.6 million molecules, of which 1.3 million are used for training. For the two datasets, we apply a preprocessing step similar to that of DiGress [24] and FreeGress [19]. Further preprocessing details and dataset statistics can be found in Appendix D.

5.1 Molecular Generation

Setup We evaluate generation quality using several metrics: Validity measures the percentage of molecules that satisfy basic valency rules. Uniqueness quantifies the proportion of molecules with distinct SMILES strings, indicating non-isomorphic structures. Novelty assesses the fraction of generated molecules absent from the training set. The filter score evaluates the proportion of molecules passing the same filters used to construct the test set. The Fréchet ChemNet Distance (FCD) compares the similarity between training and test set molecules using embeddings learned by a neural network. SNN represents the similarity to the nearest neighbor, measured via Tanimoto distance. Scaffold similarity analyzes the frequency distribution of Bemis-Murcko scaffolds, while KL divergence compares the distributions of various physicochemical descriptors.

Baselines We compare our THGD with several state-of-the-art molecular graph generative models, including JT-VAE [9] which decomposes molecules into tree-structured motifs for generation, GraphINVENT [17] which constructs molecular graphs through canonical breadth-first search ordering, NAGVAE [13] which encods substructural patterns of molecular graphs into edge features for scalable generation, MCTS [8] which is enhanced by Monte Carlo tree search for efficient chemical space exploration, as well as DiGress [24], DisCo [25] and Cometh [22] are discrete molecular graph diffusion models, operating in discrete-time, discrete-state, and continuous-time settings, respectively.

Table 1: **Molecular Generation on MOSES.** JT-VAE and GraphINVENT have hard-coded rules to ensure high validity, others do not.

Model	Validity(%) ↑	Unique(%) ↑	Novelty(%) ↑	Filter(%) ↑	FCD ↓	SNN ↑
JT-VAE [9]	*100.0*	*100.0*	*100.0*	*97.8*	*1.00*	0.53
GraphINVENT [17]	96.4	92.7	–	95.0	1.22	*0.54*
DisCo [25]	88.3	**100.0**	**97.7**	95.6	1.44	0.50
Cometh [22]	90.5	**100.0**	92.6	**99.1**	1.27	**0.54**
DiGress [24][3]	<u>96.5</u>	**100.0**	<u>95</u>	95.0	<u>1.20</u>	<u>0.51</u>
THGD	**96.8**	**100.0**	94.2	<u>98.3</u>	**1.17**	**0.54**

Table 2: **Molecular Generation on GuacaMol.** We report scores, higher is better for all metrics. NAGVAE and MCTS are tailored for molecule datasets, which incorporate more in-depth domain knowledge into the model. Others are general graph generation models.

Model	Validity (%) ↑	Unique (%) ↑	Novelty (%) ↑	KLdiv ↑	FCD ↑
NAGVAE [13]	92.9	95.5	*100*	38.4	0.90
MCTS [8]	*100*	*100*	95.4	*82.2*	*1.50*
DiGress [24]	85.2	**100**	**100**	92.9	68.00
DisCo [25]	86.6	86.6	86.5	92.6	59.7
Cometh [22]	**98.9**	<u>98.9</u>	97.6	**96.7**	<u>72.7</u>
THGD	<u>94.2</u>	**100**	<u>99.2</u>	<u>94.4</u>	**80.26**

Results Analysis THGD demonstrates strong performance in generating diverse and realistic molecules, particularly excelling in distribution fidelity, which measures how closely the generated molecules resemble real-world chemical compounds. On the GuacaMol dataset, THGD achieves a leading Fréchet ChemNet Distance (FCD) score of **80.26** among diffusion-based methods, and similarly obtains the best FCD of **1.17** on MOSES. Lower FCD scores signify a better

match to true molecular distributions, indicating THGD's superior ability to learn and replicate these complex patterns. This is further supported by a high KL divergence score of **94.4** (on GuacaMol), showing good agreement with various physicochemical properties of known molecules.

Beyond distributional alignment, THGD consistently generates high-quality structures. It achieves near-perfect chemical validity (the ability to produce chemically correct molecules) with scores of **96.8%** on MOSES and **94.2%** on GuacaMol. Notably, this is achieved without relying on hard-coded chemical rules or domain-specific knowledge, unlike specialized models such as JT-VAE, GraphINVENT, NAGVAE, and MCTS, showcasing its robust learning capabilities. Furthermore, THGD excels in generating entirely distinct and new molecules (up to **99.2% novelty** on GuacaMol), highlighting its potential for discovering novel chemical entities.

These results validate the effectiveness of our topology-aware hierarchical framework. By integrating spectral-preserved coarsening (to capture global structure) with type-specific marginal priors (to guide local atomic details), THGD ensures chemically realistic substructure generation while mitigating common distributional biases, leading to a better overall generation quality.

5.2 Scaffold-Constrained Generation

Setup In this section, we evaluate THGD's scaffold-constrained generation capabilities using four diverse scaffold templates, including three medium-sized scaffolds and one large scaffold consisting of 25 atoms, as described by Maziarz et al. [16]. The model is constrained to generate molecules that are at least as large as the given scaffold. Each scaffold was sampled four times with a batch size of 512.

Results Analysis Table 3 clearly demonstrates THGD's superiority in scaffold-constrained generation. Our model is the first graph diffusion model capable of handling large scaffolds while maintaining high validity. THGD consistently achieves impressive performance with high validity over 88% across all samples, while DiGress struggles with increasingly large scaffolds and fails with the Sildenafil case, underscoring the limitations of atomic-level diffusion in managing long-range dependencies. We further show more examples in Fig.5.

The key to THGD's success lies in its hierarchical approach, which encodes complex scaffolds into compact coarse graphs with drastically fewer nodes. For example, the modified Sildenafil scaffold, which consists of 25 atoms, is efficiently represented by only 5 coarse nodes, as shown in Fig. 6. This reduction in complexity enables more manageable refinement while preserving scaffold integrity. THGD compactly encodes scaffolds and confines refinement to peripheral regions by adopting a vocabulary-free hierarchical paradigm, ensuring precise structural control and enabling flexible molecular modifications. These capabilities are

Table 3: **Scaffold-Constrained Generation**. DiGress serves as a baseline, evaluated on the three smaller scaffolds using its publicly available source code and provided checkpoint.

Scaffold	**DiGress**	**THGD**	Improvement(%)
1,4-Dihydroquinoline[4]	46.8	**93.0**	98.7
1,3-Dimethylquinolin-4(1H)-one[5]	35.9	**88.7**	147.0
3-(Trifluoromethyl) aniline[6]	32.3	**94.7**	193.2
Modified-Sildenafil[7]	\	**88.6**	-

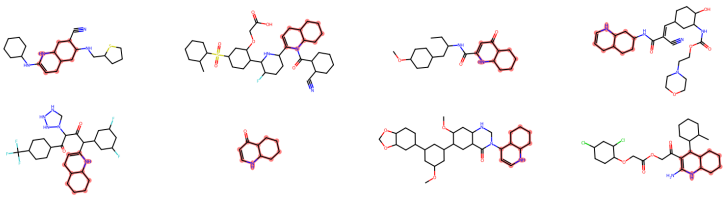

Fig. 5: Scaffold constrained generation result of 1,3-Dimethylquinolin-4(1H)-one.

particularly critical in drug discovery, where maintaining scaffold integrity while exploring chemical diversity is essential for optimizing therapeutic candidates.

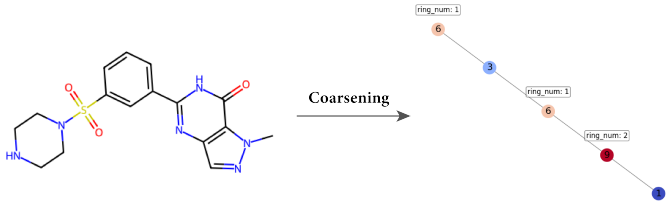

Fig. 6: Modified-Sildenafil and its coarse graph. The scaffold containing 25 atoms, is efficiently represented by only 5 coarse nodes after coarsening.

[3] DiGress is re-run here using the same settings as in the original paper after dataset preprocessing.

[4] `https://pubchem.ncbi.nlm.nih.gov/compound/1_4-Dihydroquinoline`

[5] `https://pubchem.ncbi.nlm.nih.gov/compound/3-_Trifluoromethyl_aniline`

[6] `https://pubchem.ncbi.nlm.nih.gov/compound/1_3-Dimethylquinolin-4_1H_-one`

[7] `https://pubchem.ncbi.nlm.nih.gov/compound/Sildenafil`

6 Conclusion

In this work, we introduced THGD, a Topology-Aware Hierarchical Graph Diffusion Model designed to address the challenges of generating large, structurally complex molecules. Our model innovatively decouples global topology preservation and local atomic refinement through a coarse-to-fine framework, enabling scalable and fine-grained control over molecular graph generation. By leveraging spectral-preserved graph coarsening and type-specific marginal priors, THGD eliminates reliance on predefined motif vocabularies. Experimental results on MOSES and GuacaMol benchmarks demonstrate SOTA performance, achieving an FCD score of 80.26 on GuacaMol and up to 2× higher scaffold validity than previous SOTA model in scaffold-constrained generation tasks. THGD's success highlights the potential of hierarchical diffusion models to revolutionize de novo molecular design, paving the way for more efficient drug discovery pipelines.

Acknowledgments. Thanks to the support of the Super Computing Center of Lanzhou University and Professor Yang Zhang's GPU clusters.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Anderson, W.J.: Continuous-time Markov chains: An applications-oriented approach. Springer Science & Business Media (2012)
2. Brown, N., Fiscato, M., Segler, M.H., Vaucher, A.C.: Guacamol: benchmarking models for de novo molecular design. Journal of chemical information and modeling **59**(3), 1096–1108 (2019)
3. Dai, H., Tian, Y., Dai, B., Skiena, S., Song, L.: Syntax-directed variational autoencoder for structured data. arXiv preprint arXiv:1802.08786 (2018)
4. De Cao, N., Kipf, T.: Molgan: An implicit generative model for small molecular graphs. arXiv preprint arXiv:1805.11973 (2018)
5. Gómez-Bombarelli, R., Wei, J.N., Duvenaud, D., Hernández-Lobato, J.M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T.D., Adams, R.P., Aspuru-Guzik, A.: Automatic chemical design using a data-driven continuous representation of molecules. ACS central science **4**(2), 268–276 (2018)
6. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
7. Hoogeboom, E., Satorras, V.G., Vignac, C., Welling, M.: Equivariant diffusion for molecule generation in 3d. In: International conference on machine learning. pp. 8867–8887. PMLR (2022)
8. Jensen, J.H.: A graph-based genetic algorithm and generative model/monte carlo tree search for the exploration of chemical space. Chemical science **10**(12), 3567–3572 (2019)
9. Jin, W., Barzilay, R., Jaakkola, T.: Junction tree variational autoencoder for molecular graph generation. In: International conference on machine learning. pp. 2323–2332. PMLR (2018)

10. Jin, W., Barzilay, R., Jaakkola, T.: Hierarchical generation of molecular graphs using structural motifs. In: International conference on machine learning. pp. 4839–4848. PMLR (2020)
11. Jo, J., Lee, S., Hwang, S.J.: Score-based generative modeling of graphs via the system of stochastic differential equations. In: International conference on machine learning. pp. 10362–10383. PMLR (2022)
12. Kusner, M.J., Paige, B., Hernández-Lobato, J.M.: Grammar variational autoencoder. In: International conference on machine learning. pp. 1945–1954. PMLR (2017)
13. Kwon, Y., Lee, D., Choi, Y.S., Shin, K., Kang, S.: Compressed graph representation for scalable molecular graph generation. Journal of Cheminformatics **12**, 1–8 (2020)
14. Loukas, A.: Graph reduction with spectral and cut guarantees. Journal of Machine Learning Research **20**(116), 1–42 (2019)
15. Ma, C., Zhang, X.: Gf-vae: a flow-based variational autoencoder for molecule generation. In: Proceedings of the 30th ACM international conference on information & knowledge management. pp. 1181–1190 (2021)
16. Maziarz, K., Jackson-Flux, H., Cameron, P., Sirockin, F., Schneider, N., Stiefl, N., Segler, M., Brockschmidt, M.: Learning to extend molecular scaffolds with structural motifs. arXiv preprint arXiv:2103.03864 (2021)
17. Mercado, R., Rastemo, T., Lindelöf, E., Klambauer, G., Engkvist, O., Chen, H., Bjerrum, E.J.: Graph networks for molecular design. Machine Learning: Science and Technology **2**(2), 025023 (2021)
18. Mullard, A., et al.: The drug-maker's guide to the galaxy. Nature **549**(7673), 445–447 (2017)
19. Ninniri, M., Podda, M., Bacciu, D.: Classifier-free graph diffusion for molecular property targeting. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 318–335. Springer (2024)
20. Polykovskiy, D., Zhebrak, A., Sanchez-Lengeling, B., Golovanov, S., Tatanov, O., Belyaev, S., Kurbanov, R., Artamonov, A., Aladinskiy, V., Veselov, M., et al.: Molecular sets (moses): a benchmarking platform for molecular generation models. Frontiers in pharmacology **11**, 565644 (2020)
21. Simonovsky, M., Komodakis, N.: Graphvae: Towards generation of small graphs using variational autoencoders. In: International Conference on Artificial Neural Networks. pp. 412–422. Springer (2018)
22. Siraudin, A., Malliaros, F.D., Morris, C.: Cometh: A continuous-time discrete-state graph diffusion model. arXiv preprint arXiv:2406.06449 (2024)
23. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International conference on machine learning. pp. 2256–2265. PMLR (2015)
24. Vignac, C., Krawczuk, I., Siraudin, A., Wang, B., Cevher, V., Frossard, P.: Digress: Discrete denoising diffusion for graph generation. arXiv preprint arXiv:2209.14734 (2022)
25. Xu, Z., Qiu, R., Chen, Y., Chen, H., Fan, X., Pan, M., Zeng, Z., Das, M., Tong, H.: Discrete-state continuous-time diffusion for graph generation. arXiv preprint arXiv:2405.11416 (2024)
26. Yang, M., Yang, Z., Ning, Z., Shen, H., Mao, C., Ma, C., Hu, B.: Echoes of empathy: A symbiotic iot-based emotion feedback framework for psychological interventions via large language model. IEEE Internet of Things Journal (2025)
27. Zang, C., Wang, F.: Moflow: an invertible flow model for generating molecular graphs. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 617–626 (2020)