Knowledge and Information Systems
https://doi.org/10.1007/s10115-025-02610-3

RESEARCH

Soft causal learning for generalized molecule property prediction: An environment modeling perspective

Limin Li[1] · Zhengyang Zhou[1,2] · Kuo Yang[2] · Wenjie Du[1,2] · Pengkun Wang[1,2] · Yang Wang[1,2]

Received: 25 June 2025 / Revised: 10 September 2025 / Accepted: 25 September 2025

Abstract

Learning on molecule graphs has become an increasingly important topic in AI for science, which takes full advantage of AI to facilitate scientific discovery. Existing solutions on modeling molecules utilize Graph Neural Networks (GNNs) to achieve representations, but they mostly fail to adapt models to out-of-distribution (OOD) samples. Although recent advances on OOD-oriented graph learning have discovered the invariant rationale on graphs, they still ignore three important issues, i.e., 1) the expanding atom patterns regarding environments on graphs lead to failures of invariant rationale-based models, 2) the associations between discovered molecular subgraphs and corresponding properties are complex where causal substructures cannot fully interpret the labels, and 3) the interactions between environments and invariances can influence with each other and thus are challenging to be modeled. To this end, we propose a soft causal learning framework, to tackle the unresolved molecule OOD challenge, from the perspective of negatively modeling the molecule environments and bypassing the invariant subgraphs. Specifically, we first incorporate chemistry theories into our graph growth generator to imitate expanded environments and then devise a GIB-based objective to disentangle environment from whole graphs and finally introduce a cross-attention-based soft causal interaction, which allows dynamic interactions between environments and invariances. We perform extensive experiments on seven datasets by imitating different kinds of OOD

✉ Zhengyang Zhou
zzy0929@ustc.edu.cn

Limin Li
lilimin@mail.ustc.edu.cn

Kuo Yang
yangkuo@mail.ustc.edu.cn

Wenjie Du
duwenjie@ustc.edu.cn

Pengkun Wang
pengkun@ustc.edu.cn

Yang Wang
angyan@ustc.edu.cn

[1] School of Software Engineering, University of Science and Technology of China, Jinzhai Road, Hefei 230000, Anhui, China

[2] Suzhou Institute for Advanced Research, University of Science and Technology of China, Renai Road, Suzhou 215000, Jiangsu, China

Published online: 08 October 2025

generalization scenarios. Extensive comparison, ablation experiments as well as visualized case studies demonstrate well generalization ability of our proposal.

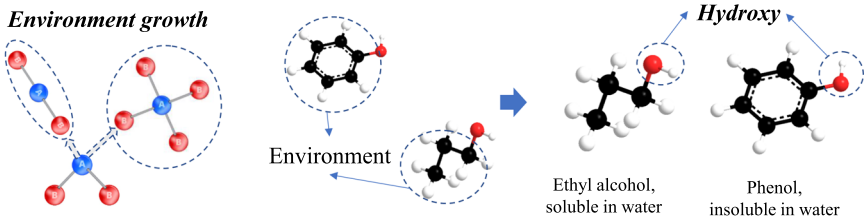
Keywords Graph neural network · AI for Science · Molecule science · Out-of-distribution generalization

Learning on molecules has increasingly become a powerful research topic to enable various applications from molecular property estimation [1], drug discovery to molecule retrosynthesis [2, 3], hence benefiting the community of scientific computing [4, 5]. However, molecule properties are mostly tested by labor-intensive experiments with the risk of poisonousness, while the drug discovery process usually costs numerous trial and errors. To this end, how to ensure the efficiency of implementing both academia and industry experiments and maximumly exploiting the power of data intelligence for practical biology and chemistry production become the central attention of researchers.

There have been numerous efforts of various Graph Neural Networks (GNNs). Technically, graph learning on molecular science either focus on finding support invariant substructures for property estimation [6–9], or exploring the homophily and heterophily on graphs to improve the representation capacity for final classification [10]. However, given the explosions of emerging materials [11] and diversity of biological medicine [12], molecular science suffers the inherent insufficiency of the training sets for model learning. Therefore, generalizing learning models for molecular graph property prediction becomes the core obstacle for GNN-based molecular research toward material-oriented industrial practices and further advances.

Recent efforts have been made to construct a series of OOD learning frameworks [7, 9, 13–15]. These solutions can be divided into three aspects, i.e., finding invariant substructure rationales [9], counterfactual-based data augmentation [14] as well as the environment augmentations [16–18]. Specifically, pioneering research devises a dual optimization strategy, which allows the joint condensation on content subgraph and neural structures [19].

Indeed, a graph is widely acknowledged that it can be decomposed into environments and causal invariant rationales. However, with the scale and diversity of molecular graphs increasing, there are two significant issues hindering the generalization of existing OOD solutions [7, 9, 14, 19, 20], as shown in Fig. 1. First, the increasing types of molecules lead to expanding patterns of environments on graphs, as illustrated in Fig. 1a where the environmental information can dominate the entire graph, resulting in the failure of invariant rationale-based models. Second, the associations between discovered molecular subgraphs and corresponding properties (i.e., labels) are complex. Given a molecule graph with specific



(a) Illustration of increasing environment on graphs.

(b) Illustration of different contexts

(c) Chemical substances with the same functional group but different properties

Fig. 1 Motivation of CauEMO. **a** Increasingly growing environments can finally dominate the property of whole graph. **b** Two substances, ethyl alcohol and phenol, are with the same functional group of hydroxyl, but are with different environmental substructures connection, resulting in different solubility properties

functional groups, the deterministic property (e.g., solubility) is not fully dependent on the specific functional groups but may partially rely on the environmental substructures around the functional group, which is demonstrated in Fig. 1b-c. From the perspective of information theory, the information encapsulated in invariant subgraphs is incomplete to interpret labels, and the reason lies in that environments usually have interactive effects with invariant rationales. Therefore, only modeling invariant rationales leads the model trapping into suboptimal results. A potential solution toward more powerful generalization capacity is to maximize the informativeness by exploiting the auxiliary environments and coordinating between environment and invariance substructures. Unfortunately, most existing works focus on subgraph extraction from invariance perspective [8, 9], while very few literature reports to decouple environmental parts and dissects how to couple extracted invariance with environmental substructures. Given the expanding molecules and unlimited synthetic materials, boosting the representation of graph environments and model-aware informativeness for generalization improvement is still faced with two specific challenges.

- The environment patterns are diverse and variable; thus, it is challenging to separate and imitate, especially how to imitate the expansion and diversity environments with theoretical guarantee.
- With environments and label-interpreted part well separated, how to cooperatively exploit the environment and invariant signals to achieve diverse property predictions?

To address the above two challenges, we propose a Molecular Property prediction network named CauEMO, to systematically tackle the OOD issue. Firstly, to promote the diversity of environments on graphs thus alleviating the potential dominance of environments in predictions, we design a knowledge-enhanced environment growth generator to simulate the environments for diverse expansions. Secondly, to improve quality of environmental representation, we treat the environment as a mediating variable and explicitly extract such representation via constructing an Environment-Graph Information Bottleneck learning objective to disentangle label-irrelevant environmental signals and label-relevant signals, allowing sufficient environment squashing. Lastly, given the potential interactions between rationales and environments and limited interpretability between rationales and labels, it is difficult to obtain a complete view of labels solely relying on invariant information. We then design an Environment-Invariance Soft Causal Interaction, which refines environment and allows information interactions between environments and causal invariance. An environment-invariance cross-attention is introduced to realize adaptive information fusion with soft scores dependent on input features. **The contributions of our work can be fourfold.**

- We discover two main factors constraining molecular graph generalization capacity. The first is that the ever-growing and expanding environmental signals on graphs gradually suppress primary information, and the second is the potential interactions between causal invariance and environments. We then propose an environment-oriented solution to increase graph diversity and capture environment-rationale interactions to enhance graph representation.
- Technically, bypassing the exploration of invariant subgraphs, we start the research from negative environmental modeling. We first explicitly incorporate chemistry principle into our graph growth generator to imitate the environment expansions and introduce a cross-attention soft causal interaction, which allows flexible and dynamic interactions between environments and invariances.
- Empirically, we conduct experiments on seven datasets, including two categories of DrugOOD datasets and one synthetic Motif dataset. These experiments demonstrate the

effectiveness of CauEMO, and the practical capacity on generalizing models to unseen graphs with increased environments and designed neural architecture.
- Our CauEMO illustrates that data-driven, AI-based approaches, specifically a framework grounded in causal invariance and invariant-environment interactions, can enhance both efficiency and interpretability in molecular property prediction, which accelerates the chemical research and support applications in drug discovery and material design.

1 Related Work

Graph neural network and subgraph learning. Graph neural networks (GNNs) are initially introduced by [21] for graph-structured data mining by iteratively aggregating information from neighbors. Recently, with the increasing prosperity of deep learning, GNNs have developed by stacking layers and simplifying the node-level adjacencies to gain powerful representation capacity, such as GCN [22], GraphSAGE [23], and Graph Attention Networks (GAT) [24], where GAT allows flexible node-level attention. However, conventional deep GNNs usually lack interpretability to explain which specific substructure contributes most to final predictions. To this end, subgraph learning is leveraged to boost the interpretability and generalization. Specifically, SubGNN [25] decouples the graph topology into three property-aware channels to extract subgraph patterns on position, neighborhood, and structure. Furthermore, Sugar [26] and XGNN [27] devise the reinforcement learning to help extract interpretable subgraphs, and P2GNN [8] is proposed to extract the asymmetric patterns of substructures in large-scale graphs with considering pivot nodes. However, generalizing models to other unseen scenarios requires capturing invariance across scenarios. Even so, these solutions to graph and subgraph learning fail to explicitly involve invariant factors thus trapped into suboptimal results in most generalization tasks.

Invariant learning for OOD generalization. Generalization issues are common in learning-based solutions, ranging from computer vision [28, 29], static graph learning [9, 30, 31] to dynamic graph learning [32, 33]. Among them, existing solutions to graph out-of-distribution generalization usually divide the whole graph into environments and invariant rationales [9], which is inherited from causal theory [34]. Representative invariant learning frameworks [30, 35] have been proposed to handle distribution shifts where it minimizes the summarized risks across different conditions and environments. Following it [30], graph-level learning for OOD generalization such as DIR [9], OOD-GNN [31], and MoleOOD [36] has been proposed for molecular scientific research. And an OOD solution on dynamic graph learning CauSTG is devised to capture invariance across sample groups [33]. Recently, EERM [37] overcomes the non-i.d.d. issue on node-level learning and takes a reinforcement learning to enhance the environment diversity. Even though, all these solutions only focus on the invariant factors, directly ignoring the valuable information within environments. Actually, graph environments can be deemed as the conditions to invariances, where environments are also equipped with valuable information and can potentially interact with invariances to influence label interpretation. Therefore, how to exploit environments to enhance cooperative learning on both environment and invariant factors for better generalization still remains under-explored.

Environment-aware learning for graph OOD generalization. Modeling the environments such as environment representation, generation [17, 38], and augmentation [14, 18] can be another way to promote OOD capacity. For instance, Zhao, et, al. consider the graph topology

as the virtual environment and devise a counterfactual strategy by imposing perturbation on environments, which can be viewed as a data augmentation [14]. CaST employs the backdoor adjustment by a novel disentanglement block to separate the temporal environments via structural causal model [18]. Moreover, [16] imposes an environment augmentation by introducing an assistant model by maximizing the variations to handle the OOD issue. Besides, [17] designs a sampling-generative process to generate new environments, while [38] disentangles the environment representation and imposes an environment-aware contrastive representation learning. Even though, above solutions either construct the closed environment set, or devise IRM and contrastive learning-based strategies to disentangle environments. In fact, the types of molecules are usually becoming more and more diverse and the number of atoms are increasing. Then, the molecules as well as corresponding graph environments cannot be fully enumerated where the limited extracted invariance cannot fully reflect the summarized properties. Hence, these above-mentioned solutions fail to mimic the increasing growths of spurious substructures and cannot capture the environment-invariance interactions.

Summary. Considering abovementioned graph learning frameworks for molecular science, there are still two significant issues that remain unresolved in OOD tasks, i.e.,

- The increasing types and numbers of molecules lead to expanding patterns of environments and result in the failure of invariant rationale-based models.
- The associations between molecular graphs and corresponding properties are complex, while interactions between environment-invariance are intractable to capture. To this end, detouring the invariance and directly enhancing environment modeling can be a promising avenue toward OOD learning improvement.

2 Preliminary and problem definition

Consider a molecule graph $G = (\mathcal{V}, \mathcal{E})$ where the node and edge in G can be denoted as $v_i \in \mathcal{V}$ and $e_{ij} \in \mathcal{E}$. The deterministic observation in node v_i is written as $x_i \in \boldsymbol{X}$, where it shows the representation of atom. Given a graph G_j, the molecular science learning is to predict a series of property $\boldsymbol{y}_j$, which consists of both graph-level regression for continuous property and graph-level classification for categorical properties.

OOD settings. Given a series of molecular graphs in training set $(\boldsymbol{Y}_{tr}, \mathcal{G}_{tr})$ and testing set $(\boldsymbol{Y}_{test}, \mathcal{G}_{test})$ where $P(\mathcal{G}_{tr}) \neq P(\mathcal{G}_{test})$, we are going to derive a neural function $y = f^*(G)$ with OOD learning capacity that can transfer invariance and adapt new environments to new scenarios.

3 Methodology

3.1 Framework overview

As shown in Fig. 2, the proposed CauEMO is composed of three well-designed components, i.e., a Knowledge-enhanced environment generator, an E-GIB for irrelevant environment disentanglement and an Environment-Invariance Soft Causal Interaction (SCI), to, respectively, imitate the increasing expansions of environments on molecule graphs with chemical knowledge constraints, disentangle environment information from whole graphs with information

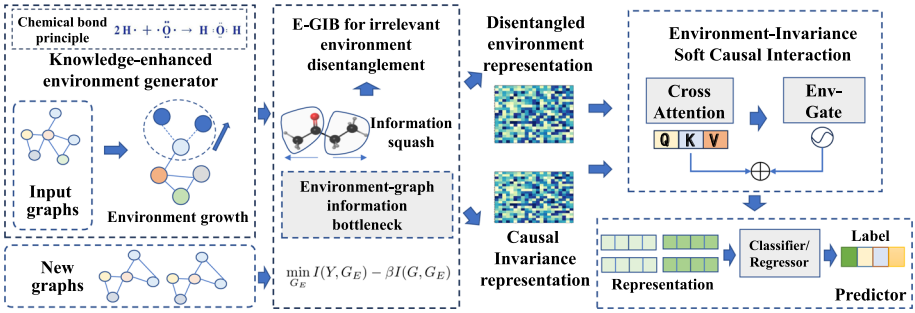


Fig. 2 Framework overview of CauEMO

theory guarantee, and dynamically implement environment refinement and environment-invariance interactions for OOD-oriented representation improvement.

3.2 Knowledge-enhanced environment generator

The environment, separated from core property substructures, within a molecular graph is also vital for property forecasting, especially for Out-of-Distribution scenarios. The diversity of graph environments can determine the accuracy and epistemic uncertainty of substructure extraction. Unfortunately, existing OOD learning either exploit the disentangled environments within the dataset itself [39], or explore an augmentation without any constraints of domain knowledge [17]. In contrast, in our paper, we devise an environment generator with the help of domain knowledge, to imitate the growth of environment over molecular graphs and simultaneously maintain the primary principle of chemistry. We first introduce the chemical knowledge-based grouping strategy to decompose the chemical unit into functional group set $\mathcal{G}_I = \{G_i\}$ and non-deterministic group set $\mathcal{G}_N = \{G_n\}$, where the former ones can be seen as the causal invariance for labels, while the latter ones are environment sets. In this way, we can dynamically combine the causal invariance with the environment part to increase the environment of graphs by generating substantial new graphs. To realize it, we rank the substructures in G_n by the number of atoms for imitation of increasing scales of environments, i.e., EgoGraph $= \{s_1, s_2, \ldots, s_m\}$, we can iteratively replace the s_k in the combined graph. However, regarding graph editing tasks, a serious issue is how to guarantee the rationality of new graphs, i.e., how to ensure the new graphs to satisfy the required chemistry properties, and then it can be synthesized for real-world industrial chemical engineering in advanced material or other fields. Hence, we introduce the chemical crosslinks to help judge whether the substructure connection between two set items are reasonable, where the domain-specific knowledge is interpreted as the law of conservation of charge. Given the connected atom v_i from $\mathcal{G}_I$, we check the summation of its chemical crosslinks. Let us denote the chemical bond of v_i as d_i, v_j are a series of neighboring nodes which is going to connected to v_i in set $\mathcal{G}_N$, then we can derive the equation according to the principle of chemical bond, i.e.,

$$d_i = \sum_j \in N(v_i) p_{ij} \Leftrightarrow v_i - v_j \tag{1}$$

If such combination $v_i \in \mathcal{G}_I, v_j \in \mathcal{G}_N$ can satisfy Equation. (1), we can make the concatenation on the node-level to achieve the new graph, i.e.,

$$G_0 = G_i, \quad G_{new} = \text{Concat}[G_0; G_j] \tag{2}$$

Instead of constructing an environment set from closed training samples without introducing any new molecules [17, 18], we especially inherit the chemical domain knowledge and concatenate the causal invariant part with auxiliary environment part iteratively. To this end, our environment growth generator can increase the uncertainty and diversity of molecules with limited training data and simultaneously ensure fundamental chemistry-specific principle for generated new samples. We believe this environment generator can work cooperatively with following learning modules to benefit downstream OOD tasks.

3.3 E-GIB-based irrelevant environment disentanglement

Most solutions to graph OOD challenges emphasize the causal invariance for transfer. But unfortunately, in real-world molecule-oriented tasks, the invariance across all graphs is limited; in other words, there is limited common invariant parts across all scenarios that can sufficiently support the transfer. In our work, we focus more on the environments, which can be further refined to enhance the improvement and informativeness of causal parts. We thus bypass modeling invariant associations and propose an Environment-Graph Information Bottleneck (E-GIB) to explicitly extract the environments that are mostly irrelevant with deterministic graph labels, from the perspective of information theory.

Given a graph G, the initialized environment representation G_E, and the label Y of G, our E-GIB is expected to find out the most label-irrelevant representation on graph G_E. Our E-GIB will not only squash the environment representation away from labels but also ensures the environment G_E can cover most of the graph G. By borrowing the theoretical guarantee from information theory, the guided training objective can be preliminarily described as,

$$\min_{G_E} I(Y, G_E) - \beta I(G, G_E) \tag{3}$$

where β is the hyperparameter setting as 1 according to common practices [40–42]. We will then elaborate the implementation of above E-GIB.

For the first term in Equation. (3), we take the standard cross-entropy loss to instantiate the preliminary objective, which aims to suppress the label-relevant information on graphs by inheriting the tractable lower bound obtained from literature [42]. This learning objective is specified as the environment predictor $P_\theta(Y|G_E)$ (a.k.a. f_θ). Regarding the second term, since there is no premise or apriori information for marginal distribution $p(G_E)$, we derive a variational estimator $\mathbb{Q}(G_E)$ to approximate $p(G_E)$, i.e., $p(G_E) \sim \mathbb{Q}(G_E)$, and obtain the variational upper bound with KL-divergence [43]. It can be formally derived by,

$$I(G, G_E) \leq \mathrm{KL}\left(P_\phi(G|G_E)||\mathbb{Q}\left(G_E\right)\right) \tag{4}$$

Then, we can take such KL-divergency to estimate the marginal probability $p(G_E)$ with parameterized $\mathbb{Q}(G_E; \boldsymbol{W}_{ge})$, where $\boldsymbol{W}_{ge}$ are learnable parameters to this probability distribution estimation. We can designate the P_ϕ (a.k.a. g_ϕ) as the environment extractor.

Learning objective. Considering both cross-entropy for label-irrelevant information suppression and the KL-divergence for variational estimator, we can obtain the final learning objective for our environment disentanglement, i.e.,

$$\min \ \mathbb{E}[log\mathbb{P}_\theta(Y|G_E)] - \beta\mathbb{E}[\mathrm{KL}(P_\phi(G|G_E)||\mathbb{Q}(G_E)] \tag{5}$$

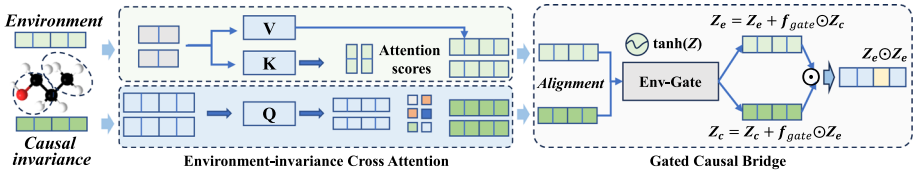


Fig. 3 Environment-invariance soft causal interaction

With above learning objective, we can sufficiently disentangle the environment G_E over graph G. Then, we can provide the details of how to implement our environment extractor and environment predictor.

Implementation of environment extractor g_ϕ. The environment extractor g_ϕ encodes input graph G via GNN and for each edge $(u, v) \in \mathcal{E}$, g_ϕ consists of an MLP layer and a sigmoid function that maps the concatenation of two node representation into $p_{uv} \in [0, 1]$, i.e.,

$$p_{uv} = \text{MLP}(h_u, h_v; \boldsymbol{W}_g) \tag{6}$$

where (h_u, h_v) are representations of node u and v, $\boldsymbol{W}_g$ are learnable parameters for environment extractor. For each forward pass during training process, we sample stochastic attention from Bernoulli distributions $\alpha_{u,v} \sim \text{Bern}(p_{uv})$. We also apply the gumbel-softmax reparameterization trick to ensure the continuous gradient for computable p_{uv} [44]. Then, the extracted graph G will have an attention-based selected subgraph as,

$$\boldsymbol{A}_S = \alpha \cdot \boldsymbol{A} \tag{7}$$

Therefore, the environment subgraph extractor falls into a Stochastic Attention mechanism controlled by Bernoulli distribution.

Implementation of environment predictor f_θ. The predictor f_θ adopts the same GNN to encode the extracted graph G to a graph representation and finally passes such representation through an MLP layer plus softmax to model the distribution of Y. This procedure enables the variational distribution $P_\theta(Y|G_E)$.

Finally, the Marginal Distribution can be controlled via $\mathbb{Q}$. Then, we can simultaneously obtain the maximal label-irrelevant information G_E and achieve the complementary subgraph to G_E, i.e., $G_I = G_E^-$, which allows further refinement and environment-invariance interactions.

3.4 Environment-invariance soft causal interaction

Motivation. It is observed that the only invariance across environments cannot sufficiently contribute to final property as some environments can be taken as the conditions and account for some molecular properties, which is illustrated in Fig. 1. Thus only exploiting the invariance for generalization is limited in information loss. In this subsection, we propose to refine the environment and enable interactions between extracted environment and invariance. Specifically, we argue that the environment-invariance interactions should include three crucial technical issues,

- Ensuring sufficient mutual interactions between environment and causal invariant substructures.
- The interpretability during learning and aggregation process.

- The dimension alignment for easy-to-implement interaction.

Solution. To systematically address above issues, we propose our Soft Causal Interaction (SCI) scheme. In order to capture the interactions between environment representation and invariant causal representation on graph, we allow partial associated environment aggregating with causal invariance in a learnable manner in our SCA. As shown in Fig. 3, our SCA consists of two parts, i.e., an Environment-invariance Cross-Attention to capture the potential correlations within environment itself and then between environment and invariance. And a Gated Causal Bridge is designed to dynamically allow the sufficient information injection and interactions between invariances and associated environment to boost the transfer capacity.

Algorithm 1 The training process of CauEMO

Input: graph dataset $\mathcal{G}$
Initial: Functional group set $\mathcal{G}_I = \{G_i\}$, non-deterministic group set $\mathcal{G}_N = \{G_n\}$, the number of epochs K.
for $i = 1$ **to** K **do**
  Knowledge-enhanced environment generator:
  $\mathcal{G} \leftarrow \{G_{new}\} = \text{Concat}[G_0; G_j]$**, where** $G_j \in \mathcal{G}_N$
  $p_{uv} = \text{MLP}(h_u, h_v; \boldsymbol{W}_g)$
  $\alpha_{u,v} \sim \text{Bern}(p_{uv})$**,** $G_E \sim \boldsymbol{A}_S = \alpha \cdot \boldsymbol{A}$
  E-GIB environment disentanglement:
  $\boldsymbol{Z}_e = \mathbb{P}_\theta(G_E)$**,** $\boldsymbol{Z}_c = \mathbb{P}_\theta(G_{\overline{E}})$
  $\boldsymbol{Z}_e^Q = \boldsymbol{Z}_e \boldsymbol{W}^Q,\ \boldsymbol{Z}_c^K = \boldsymbol{Z}_c \boldsymbol{W}^K,\ Z_c^V = \boldsymbol{Z}_c \boldsymbol{W}^V$
  $\boldsymbol{Z}_e = \text{Softmax}(\frac{\boldsymbol{Z}_e^Q (\boldsymbol{Z}_c^K)^T}{\sqrt{d}}) \boldsymbol{Z}_c^V + \varepsilon$
  $\boldsymbol{Z}_c = \text{Softmax}(\frac{\boldsymbol{Z}_c^K (\boldsymbol{Z}_e^Q)^T}{\sqrt{d}}) \boldsymbol{Z}_c^V + \varepsilon$
  Implement invariance and environment interactions with Gated Causal Bridge (Equation. 11).
  $\widehat{Y} = \text{MLP}_\psi(\boldsymbol{Z}_{ce})$
  Learning optimizing:
  $\min \ -\mathbb{E}[log \mathbb{P}_\psi(Y|G)] + \mathbb{E}[log \mathbb{P}_\theta(Y|G_E)] - \beta \mathbb{E}[\text{KL}(P_\phi(G|G_E)||\mathbb{Q}(G_E)]$
end for
Return ψ, θ, and ϕ.

In detail, our Environment-invariance Cross-Attention is composed of three learnable parameters, $\boldsymbol{W}^Q$, $\boldsymbol{W}^K$, $\boldsymbol{W}^V$, we feed the representation of environment $G_E(\boldsymbol{Z}_e)$ and causal representation $\boldsymbol{Z}_c$ into the attention mechanism. The three hidden representation for soft score calculation can be derived,

$$\boldsymbol{Z}_e^Q = \boldsymbol{Z}_e \boldsymbol{W}^Q,\ \boldsymbol{Z}_c^K = \boldsymbol{Z}_c \boldsymbol{W}^K,\ Z_c^V = \boldsymbol{Z}_c \boldsymbol{W}^V \tag{8}$$

Then, we can impose the cross-attention to capture the mutual interactions thus obtaining the environment representation $\boldsymbol{Z}_e$ and causal invariance representation $\boldsymbol{Z}_c$,

$$\boldsymbol{Z}_e = \text{Softmax}(\frac{\boldsymbol{Z}_e^Q (\boldsymbol{Z}_c^K)^T}{\sqrt{d}}) \boldsymbol{Z}_c^V + \varepsilon \tag{9}$$

$$\boldsymbol{Z}_c = \text{Softmax}(\frac{\boldsymbol{Z}_c^K (\boldsymbol{Z}_e^Q)^T}{\sqrt{d}}) \boldsymbol{Z}_c^V + \varepsilon \tag{10}$$

where the random noise satisfying $\varepsilon \sim \mathcal{N}(0, \boldsymbol{I})$ is added to boost the robustness of representations. Then, the learnable coefficient $\{\text{Softmax}(\frac{\boldsymbol{Z}_e^Q (\boldsymbol{Z}_c^K)^T}{\sqrt{d}}), \text{Softmax}(\frac{\boldsymbol{Z}_c^K (\boldsymbol{Z}_e^Q)^T}{\sqrt{d}})\}$ between 0 and 1 is considered as the correlations.

By obtaining the correlation enhanced representation, we further introduce a Gated Causal Bridge to allow partial relevant environments to be aggregated with invariance substructure representations. It is followed by three steps,

- Dimension alignment between environment and invariant substructure for the generation of our gate where the dimension of $\boldsymbol{Z}_e$ is aligned to the same with $\boldsymbol{Z}_c$.
- Absorbing the mutual information to update the respective $\boldsymbol{Z}_e$ and $\boldsymbol{Z}_c$ with interpretability.
- Implementing the interactions between $\boldsymbol{Z}_e$ and $\boldsymbol{Z}_c$ for achieving final interacted representation $\boldsymbol{Z}_{ce}$.

We can formulate the above steps in following equations,

$$\begin{cases} \boldsymbol{f}_{gate} = \tanh(\boldsymbol{W}_{gate}\boldsymbol{Z}_e) \\ \boldsymbol{Z}_e = \boldsymbol{Z}_e + \boldsymbol{Z}_c \odot \boldsymbol{f}_{gate} \\ \boldsymbol{Z}_c = \boldsymbol{Z}_c + \boldsymbol{Z}_e \odot \boldsymbol{f}_{gate} \\ \boldsymbol{Z}_{ce} = \boldsymbol{Z}_c \odot \boldsymbol{f}_{gate} \end{cases} \tag{11}$$

where $\odot$ denotes element-wise product, the activation function tanh allows both positive and negative signs for the environment output, satisfying the information filtering from environment to invariance. The $\boldsymbol{f}_{gate}$ can be viewed as the squashed environment representation, and the updated $\boldsymbol{Z}_c$ can be considered as incorporating the partial relevant environment representation with a soft weighted parameter $\boldsymbol{f}_{gate}$ for final prediction. When the training set is already, it can be further extended with knowledge-based environment generation and then fed into the our neural network to obtain,

$$\widehat{Y} = \mathrm{MLP}_{\psi}(\boldsymbol{Z}_{ce}) \tag{12}$$

where ψ are parameters for final classifier or regressor. Assuming Y_{G_i} and $\widehat{Y}_{G_i}$ are the ground-truth and predicted molecule properties of given graph G_i, then the training objective is to,

$$Loss = \sum_{G_i \in \mathbb{G}_{train}} (Y_{G_i} - \widehat{Y}_{G_i})^2 \tag{13}$$

3.5 OOD prediction stage

As the out-of-distribution molecules $\mathbb{G}_{test} = \{G_{t_1}, G_{t_2}, ?\}$ come, we can feed the new graph G_{t_i} into the molecular learning framework CauEMO. Following the irrelevant environment disentanglement and environment-invariance soft causal interaction, CauEMO can efficiently disentangle the environment part $\boldsymbol{Z}_e^{t_i}$ on G_{t_i} and further boost the causal invariance into $\boldsymbol{Z}_c$ with a cross-attention and environment gate. Then, we can take the environment-enhanced causal invariance representation with well soft aggregation $\boldsymbol{Z}_{ce}$ for final prediction can be $\widehat{Y} = \mathrm{MLP}_{\psi}(\boldsymbol{Z}_{ce})$.

4 Experiment

We evaluate CauEMO using both synthetic and real-world datasets by explicitly involving distribution shifts. Both practices of causal invariance and environment-based methods are taken for comparison. Specifically, we would like to answer the following two questions via empirical experiments:

- On scenarios where environmental information dominates the graph, can our CauEMO outperform existing methods?

- When the associations between invariant rationales and labels are implicit, can CauEMO capture true causal associations?

4.1 Dataset

The datasets for evaluation are threefold. We choose totally 6 datasets, including four real-world molecular datasets, two categories of drugOOD datasets and one synthetic dataset of Motif to verify the effectiveness of CauEMO.

There are **five** real-world datasets on molecular property prediction.

- *DrugOOD datasets*, which will be exploited to evaluate our CauEMO. To evaluate the OOD performance of CauEMO, we adopt 6 sub-datasets from two categories of **DrugOOD benchmark** [45]. It focuses on the challenging real-world task of AI-aided drug affinity prediction. The distribution shift happens on different Assays, Scaffolds and molecule Sizes. In particular, **DrugOOD-lbap-core-ec50-assay, DrugOOD-lbap-core-ec50-scaffold, DrugOOD-lbap-core-ec50-size, DrugOOD-lbap-core-ki-assay, DrugOOD-lbap-core-ki-scaffold, and DrugOOD-lbap-core-ki-size** are selected.
- **Open Graph Benchmark (OGB)** [46] is a series of real, large-scale and diverse datasets which are utilized for machine learning on graphs. It covers almost all real-world tasks, including node-level, link-level and graph-level property prediction. We choose **MOL-HIV**, **BBBP** and **SIDER** to verify our method.
- **MUTAG** [47] is a binary dataset of molecular property, where nodes indicate atoms and edges denote chemical bonds. Each graph is associated with a binary label based on its mutagenic effect.

For *synthetic datasets*, we select a synthetic dataset to assiduously verify the validity and interpretability of CauEMO. **Spurious-Motif** is a synthetic dataset proposed by [9] with three graph classes. Each graph is composed of one base S and one motif C. The motif C directly determines the label of the graph. We can create Spurious-Motif datasets with different spurious correlations, which represents the degree (b) between the base S and the label. In our implementation, we let $b = 0.5, 0.7, 0.9$ for dataset generation.

4.2 Baselines

We choose three categories of baselines, including conventional GNN backbones, subgraph-based invariant learning methods and environment-based graph learning models.

Conventional backbone baselines. Three popular backbones in most practices are taken as our baselines for evaluation.

- *GCN* [22] is a vanilla Graph Convolution Neural Network via capturing the spatial adjacency for aggregating neighborhood.
- *Graph-SAGE* [23] takes a random neighbor sampling strategy to simplify the computation of information fusion and it allows inductive learning on new nodes.
- *GIN* [48] is an isomorphism graph network to ensure the consistent structure to be with similar representations.

Subgraph-based invariant learning methods. We adopt four typical subgraph-based learning baselines for evaluation.

- *SUN* [49] studies the characteristics of node-based subgraph learning and aligns the permutation group of nodes and subgraphs, modeling the symmetry with a smaller single permutation group.

- *IB-subgraph* [42] first implements the information bottleneck with graph learning, which is not only a subgraph learning based on partition (edge drop) but also an important exploration of interpretability.
- *GSAT* [50] follows this practice and designs a subgraph extraction strategy with edge deletions based on stochastic attention mechanism.
- *DIR* [9] splits the input graph into causal and non-causal subgraphs and utilizes invariant features to construct interpretable model.

Environment-based graph learning models. We exploit five baselines with explicitly considering the modeling of graph environments. All of them focus on the graph-level out-of-distribution generalization from the perspective of environment modeling.

- *CIGA* [51] is an environment-base learning architecture, which utilizes contrastive learning within the same class labels, and assume samples with the same label share invariant substructures.
- *GALA* [16] is a symmetric graph convolutional autoencoder for unsupervised graph representation learning.
- *IGM* [52] proposes to exploit the environment to augment the learning of invariance.
- *NeGo* [53] is an environment-aware solution, which emphasizes environment as negative part for inference and tackle the OOD challenge.
- *EAGLE* [54] is also a state-of-the-art environment-aware framework for OOD generalization by modeling complex coupled environments and exploiting spatiotemporal invariant patterns.

4.3 Evaluation metrics and implementation details

We employ the same metrics as the previous approach to evaluate specific dataset. For the MUTAG and Spurious-Motif datasets, we exploit **accuracy** as the evaluation metric.

For the DrugOOD and OGB datasets, we evaluate the performances using the **ROC-AUC** metric where the value of this metric is the higher, the better. We report the mean results and standard deviations across ten runs. We exploit GIN as the backbone of CauEMO, and all experiments are conducted on an NVIDIA A100-PCIE-40GB.

4.4 Performance comparison

OOD generalization performance under distribution shifts. In Table 1, we report the ROC-AUC on six distribution-shift datasets. We can clearly observe that our CauEMO consistently achieves the best performance across five datasets. This demonstrates that our environment-centered design can achieve superior performance under distribution shift scenarios. Moreover, we also have the following two observations. 1) Compared with conventional backbone GNNs, those graph learning models specially designed for OOD scenarios have better performance. Even GIN reaches the best result on Ki-Assay, it suffers severe performance fluctuations. This explicitly confirms the validity and rationality of existing invariant learning methods and environment-based models. 2) Compared with the methods based on causal invariance theory, the environment-oriented models perform better. GALA, IGM as well as NeGo and EAGLE-mole obtain sub-optimal results across all datasets, suggesting these environment associated solutions can potentially improve OOD generalization capacity with environment-oriented strategy. This serves as the practical foundation of our work on explicit environment modeling and disentanglement. Therefore, we can conclude

that our CauEMO achieves nearly best performances among all baselines, and we believe our environment-aware and invariance-environment interaction module are superior to peer models.

Prediction performance of real-world tasks and interpretability. In Table 2, we show the prediction performance of CauEMO on four real-world datasets and three synthetic datasets. The results suggest that CauEMO achieves competitive performance in real-world molecular classification tasks, reaching five best performances across seven datasets. Noted that IGM, which utilizes the cooperative mix-up strategy combining both environment and invariance parts, slightly outperforms our CauEMO on two datasets and it can potentially verify that the intuition of environment-invariance cooperation makes sense. It is worth noting that CauEMO has gained more capacity on generalization over other five datasets, which may be attributed to the sufficient mutual interactions in environment-invariance SCA, and label-irrelevant information squash of E-GIB. We choose the synthetic dataset to explore whether CauEMO could identify specific causal substructures. As shown in Fig.4, we present the ability of CauEMO to discover the structure of 'house' around various environments in Spurious-Motif dataset. Despite the diversity of surrounding four environments in Motif, our CauEMO can always accurately identify the invariant property substructures and we believe the potential reason behind the superior performances derived from perceiving broader various environments that is complementary to invariances.

4.5 Detailed evaluation on challenging cases

Given the real-world scenarios usually meet up with noise in labels, and some molecules may reveal conflict properties of environments and invariances, it is essential to provide more detailed evaluations on these more challenging cases and observe how our CauEMO behave under these cases, which can facilitate the effectiveness and significance of our research. In this subsection, we implement two detailed cases for evaluation, 1) the label noise and 2) environment-invariance conflicts.

Performance comparisons on challenging cases with label noise. We conducted a robustness test by injecting 10% and 20% random label noise into MUTAG datasets, i.e., inverse the label from 1 to 0 and 0 to 1 on randomly selected samples. The experimental results are shown in Tab. 4, and it reveals that CauEMO degrades 17.10% (↓) on Accuracy performances when the label noise increase from 0 to 30%, while other baselines, GIN degenerates 22.48%(↓) and GALA degenerates 21.92%(↓), indicating stronger resilience to noise.

Performance comparisons on challenging cases with environment-invariance conflicts. To facilitate the understanding the interpretability and specific superiority of our CauEMO, we also provide an example of environment-invariance conflict, where the revealed property of environment and invariant parts are conflict with each other. Specifically, we constructed a synthetic conflict sample where benzene rings (invariance) determine the true label = 1, while hydroxyl groups (environment) spuriously suggest label = 0, as shown in Fig. 8. We conduct experiment on GIN and our CauEMO. As shown, in such conflict molecules (benzene + OH), our CauEMO can present the inference result '1' while other baseline GIN outputs '0'. This verifies our CauEMO can counteract the conflict with well-designed E-GIB and soft causal interaction.

Table 1 OOD generalization performance on DrugOOD datasets (ROC-AUC)

	EC50-Assay	EC50-Sca	EC50-Size	Ki-Assay	Ki-Sca	Ki-Size
GCN	61.20± 1.60	65.3 ± 1.91	52.1 ± 2.0	83.7 ± 4.70	33.2 ± 1.81	31.6 ± 1.72
Graph- SAGE	74.8 ± 3.40	64.1 ± 2.84	52.5 ± 1.63	84.6 ± 5.32	34.8 ± 2.00	31.5 ± 2.50
GIN	75.8 ± 1.31	66.4 ± 2.00	56.2 ± 1.60	89.4 ± 5.62	39.9 ± 1.31	39.0 ± 1.60
IB- subgraph	75.31±2.06	62.91±1.67	60.57±2.03	72.41±1.23	70.67±2.30	73.65±2.34
GSAT	76.07±1.95	63.58±1.36	61.12±0.66	72.26±1.76	71.16±0.80	75.78±2.60
DIR	74.51±2.12	63.23±1.44	61.82±1.04	71.92±1.23	69.56±0.43	74.98±1.96
CIGA	75.03±2.47	65.41±1.16	64.10±1.08	73.95±2.50	71.87±3.32	74.46±2.32
GALA	77.56±2.88	66.28±0.45	64.25±1.21	77.92±2.48	73.17±0.88	77.40±2.04
IGM	77.60±2.10	65.74±1.04	63.45±1.19	75.63±2.40	73.83±1.60	76.95±2.19
NeGo	72.20± 1.60	65.30 ± 1.91	52.10 ± 2.0	83.7 ± 4.70	74.02 ± 1.81	75.13 ± 1.44
EAGLE-mole	74.45± 1.78	66.32 ± 2.10	54.55 ± 3.50	84.18 ± 3.12	71.06± 1.66	73.32 ± 1.72
CauEMO (Ours)	**78.46±1.42**	**66.91±0.62**	**65.77±1.03**	**78.12±1.81**	**74.53±1.10**	**78.21±1.92**

The best results are in **bold**, and the second best is underlined

Table 2 Performance comparison. We report ROC-AUC for MOLHIV, BBBP, and SIDER. Besides, we also report Accuracy on MUTAG and Spurious-Motif for evaluation

	MOLHIV	BBBP	SIDER	MUTAG	Spurious-Motif		
					0.5	0.7	0.9
GCN	75.52±1.61	65.34±1.94	52.12±2.03	83.75±4.74	33.22±1.82	31.68±1.73	29.61±6.23
Graph-SAGE	74.82±3.42	64.16±2.83	52.52±1.69	84.60±5.34	34.88±2.08	31.53±2.50	30.42±3.47
GIN	75.86±1.30	66.46±2.00	56.24±1.64	89.42±5.63	39.96±1.35	39.04±1.60	38.62±2.33
IB-subgraph	76.43±2.65	68.12±1.12	57.71±2.14	94.33±6.44	54.36±7.09	48.51±5.76	46.19±5.63
GSAT	76.47±1.53	69.13±2.02	59.19±1.03	96.37±2.15	52.74±4.08	49.12±3.29	44.22±5.57
DIR	76.34±1.01	69.73±1.54	58.81±1.84	96.01±2.24	58.73±2.15	43.36±1.64	39.87±0.56
CIGA	76.94±1.32	69.65±1.32	58.95±1.22	95.77±1.23	77.33±9.13	69.29±3.06	63.41±7.38
GALA	77.04±1.60	70.21±1.31	59.04±1.30	96.76±1.70	73.45±5.43	68.56±3.32	69.82±2.34
IGM	77.20±1.39	71.03±0.79	58.23±1.43	96.04±2.01	**82.36±7.39**	**78.09±5.63**	76.11±8.86
CauEMO (Ours)	**77.91±1.54**	**72.31±1.02**	**59.90±1.28**	**96.92±1.36**	81.45±6.05	76.82±2.81	**77.17±5.95**

The best results are in **bold** and the second best is underlined

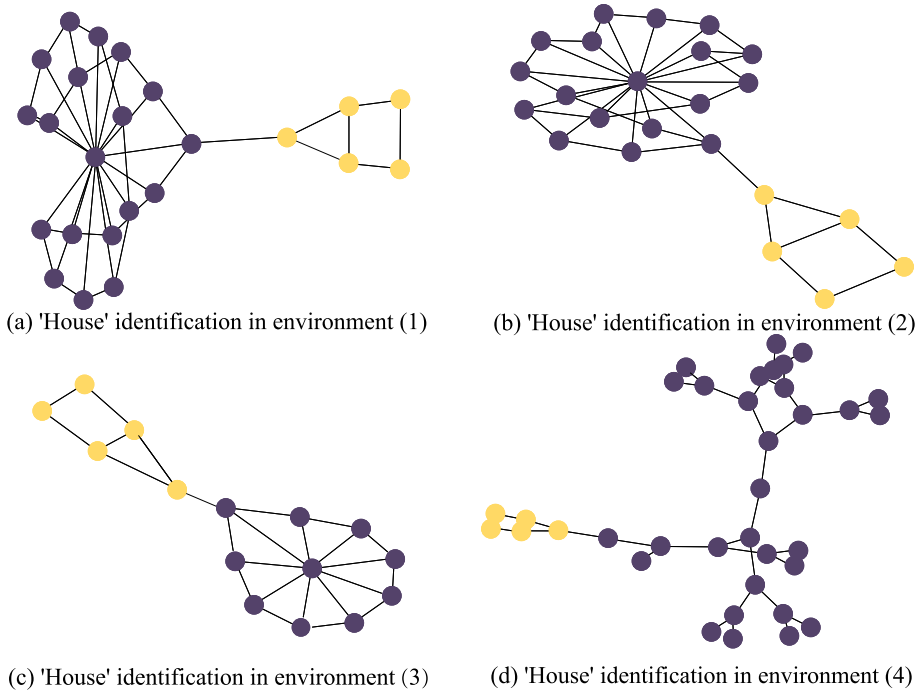


Fig. 4 The ability of CauEMO to identify 'house' in Spurious-Motif dataset

4.6 Ablation study

We conduct ablation studies from following four aspects:

- For chemistry-guided environment generator, whether the knowledge can exactly improve the generator and the OOD prediction performance, and whether our design can benefit uncover environment diversity?
- Whether the environment-centered graph information bottleneck design is superior to an invariant subgraph-centered approach?
- Whether the interaction mechanism between invariant subgraphs and environment subgraphs through cross-attention in soft causal-invariance interaction is beneficial to performance improvement?
- Whether the design of a Gated Causal Bridge is helpful for isolating environment variables while preserving the causal invariant subgraph?

Firstly, we design two variants of CauEMO, CauEMO-NonKW and CauEMO-Random, which, respectively, remove the knowledge in the generator (without chemical bond constraint) and utilize random noise to achieve environment growth. For effectiveness of knowledge utilization, we implement such ablation study on OGB (open graph benchmark) and MUTAG as well as synthetic datasets (Spurious-Motif) in Tab. 3. Without explicit molecule principle constraint, the generated molecules may deviate from the real one, that is to say, the generated molecules may not exist in real world. As observed the third-line of this table, we can find without knowledge injection, our variant degrades approximately 6%, we can speculate the reason behind it maybe the lacking of knowledge leads to abnormal molecules losing the basic regularity of chemistry and molecules thus the model cannot

Table 3 The ablative performance comparison on graph causal bridge and knowledge involvement

	MOLHIV	BBBP	SIDER	MUTAG	Spurious-Motif		
					0.5	0.7	0.9
CauEMO	**77.91**	**72.31**	**59.90**	**96.92**	**81.45**	**76.82**	**77.17**
CauEMO -NonGCB	75.52	70.14	54.02	93.37	76.58	72.41	73.62
CauEMO -NonKW	73.17	68.42	55.33	91.08	75.74	71.08	73.40

The reported are results on ROC-AUC, and the bold texts are the best results among three lines

identify the exact properties. Further, we conduct experiments of random noise injection of generator on EC50-Size, Ki-Size and MUTAG datasets. Figure 5a shows the comparison between CauEMO and CauEMO-Random. We observe that CauEMO-Random achieves a worse performance than CauEMO on most datasets. This indicates that environment variables in molecules show strong domain-specific characteristics, and we can conclude that incorporating environment information into molecules in a random and unstructured way often causes shifts in their inherent properties. Therefore, our designed chemistry-guided environment generator can effectively enhance the discovery of environmental diversity by introducing structured and contextually relevant variations, allowing for a more comprehensive exploration of diverse environmental conditions.

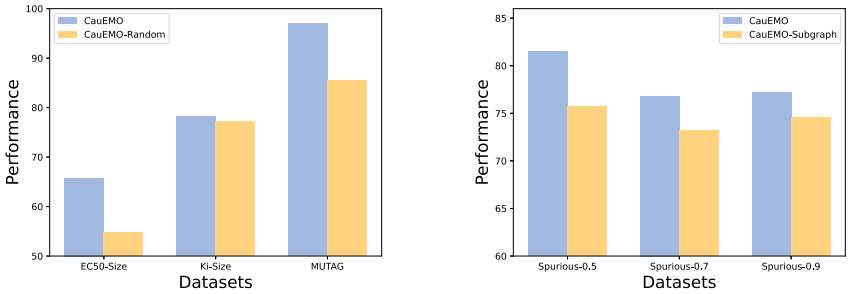
Secondly, we propose a variant centered on causal subgraph learning, CauEMO-Subgraph. This variant does not focus on modeling the environment factors but instead remains aimed at extracting causal subgraphs. On the Spurious-Motif datasets, we compare the performance of CauEMO and CauEMO-Subgraph. As shown in Fig. 5b, we observe a significant performance degradation in CauEMO-Subgraph, highlighting the effectiveness of our environment-centered approach. We argue that this performance degradation stems from the fact that positive causal learning often fails to isolate invariant subgraphs in complex environments. Modeling the environment factors, which effectively incorporate both positive and negative learning, provides a more robust solution for OOD generalization in graph learning.

Thirdly, we aim to verify whether the interaction between causal variables and environment variables in graphs can enhance model robustness. To this end, we propose a variant without an interaction mechanism, CauEMO-NonInter, where the learned environment representation does not contribute to enhancing the positive invariant learning process. As shown in Fig. 6, we can observe that the invariant subgraphs learned by CauEMO-NonInter are usually with several environment nodes those erroneously identified (Second line of Fig. 6). In contrast, our CauEMO can achieve clearer extractions of invariant subgraphs. This can potentially indicate that the interaction of causal variables and environment variables makes environment-invariance easier to separate.

Finally, we aim to validate the effectiveness of the Gated Causal Bridge. Thus, we can obtain a variant of CauEMO, CauEMO-NonGCB, by removing the Gated Causal Bridge component. As shown in Table 3, the results on all datasets reveal that the ablative variant is inferior to our CauEMO (integrated one). These empirical results further confirm that the design of Gated Causal Bridge is beneficial to enhance the effectiveness and robustness of the model.

4.7 Hyperparameter analysis

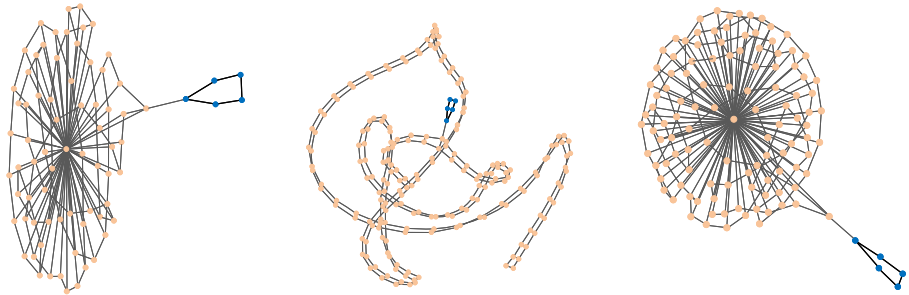
The important hyperparameters in our study are twofold. First, in E-GIB, the balance parameter β in Equation.(3) plays the role of balancing the trade-off of compositional invariance and



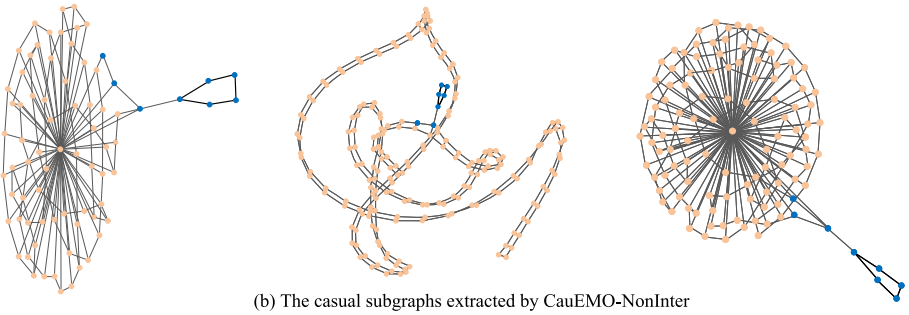
(a) Performance comparison between CauEMO and CauEMO-Random.

(b) Performance comparison between CauEMO and CauEMO-Subgraph.

Fig. 5 Ablation studies on CauEMO-Random and CauEMO-Subgraph



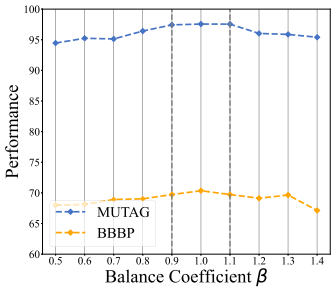
(a) The casual subgraphs extracted by CauEMO



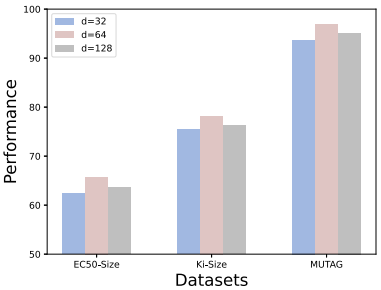
(b) The casual subgraphs extracted by CauEMO-NonInter

Fig. 6 The comparison between CauEMO and its ablative variant without interaction mechanism, CauEMO-NonInter

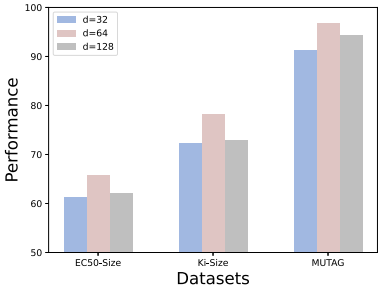
environment. We set β in the interval [0.5, 1.4] and visualize the performances on two selected real-world datasets MUTAG and BBBP for generalization task. Second, in soft causal interaction (SCI), the latent dimension of representation, i.e., molecular environment representation ($\mathbf{Z}_e$), molecular invariance representation ($\mathbf{Z}_c$) and gated environment-invariance representation ($\mathbf{Z}_{ce}$), where the dimensions are ranging in the scale $\{\mathbb{R}^{16\times1}, \mathbb{R}^{32\times1}, \mathbb{R}^{64\times1}, \mathbb{R}^{128\times1}\}$. The larger dimension may indicate more learning capacity while simultaneously means more complexity and computational workloads. We also perform such dimension-related sensitivity tests on datasets of MUTAG and BBBP on generalization task. We visualize our hyperparameter analysis process in Fig. 7. Regarding β, it experiences a climbing stage and a downward stage where it achieves best performances during the interval [0.9, 1.1] on both



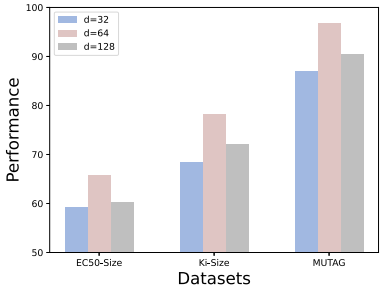
(c) Performance analysis of the hyperparameter β.



(d) Performance comparison of the dimension of $\boldsymbol{Z}_e$.



(e) Performance comparison of the dimension of $\boldsymbol{Z}_c$.



(f) Performance comparison of the dimension of $\boldsymbol{Z}_{ce}$.

Fig. 7 Performance variations during hyperparameter adjustment

two datasets, then β is set to 1 in our implementation. For the dimension of representation, it also experiences a climb and then drop down with dimension increasing, we then set 64 across all datasets as an intermediate trade-off for final experiments.

4.8 Complexity analysis

Since efficiency of artificial neural network is also important for real-world implementation and concerning about the scalability. In this subsection, we provide a thorough analysis on complexity and efficiency and further discuss the potential improvement regarding efficiency. According to our solution of stochastic attention from Bernoulli distributions, we can derive that our solution enjoys the efficiency of $\mathcal{O}(L|V|d)$ of time complexity and $\mathcal{O}(L|V|d+|E|d)$, where $|V|$ is the number of nodes, $|E|$ is the number of edge in graph G. Noted that in this scenario, we do not introduce any $|V|^2$ item that may induce time-consuming and space-consuming issue, thus there is not a large burden of computation, even for larger molecules. In addition, we empirically compare the training time efficiency of CauEMO with other baselines on EC50-Size dataset as shown in Tab. 5. Compared to traditional invariant learning approaches such as DIR and GSAT, our CauEMO achieves a substantial performance gain with only a marginal increase in runtime. Compared with state-of-the-art graph invariant learning strategies (such as CIGA, GALA, and NeGo), our CauEMO offers a significant advantage in efficiency and performance.

Table 4 OOD generalization performance on DrugOOD datasets (ROC-AUC)

	MUTAG	MUTAG-10%INV	MUTAG-20%INV	MUTAG-30%INV
GIN	75.8 ± 1.31	71.25 ± 1.65	65.42 ± 1.75	59.18 ± 3.68
GALA	77.56±2.88	68.45±2.44	64.66±1.93	60.12±2.04
CauEMO (Ours)	**78.46±1.42**	**74.15±2.41**	**68.54±1.67**	**65.12±1.81**

The best results are in **bold**

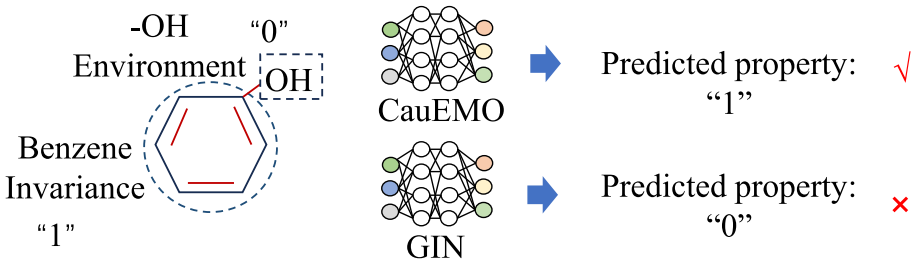


Fig. 8 A special case of environment-invariance conflict

Table 5 The training efficiency of `CauEMO` with other baselines on EC50-Size (s/epoch)

Models	GSAT	DIR	CIGA	GALA	NeGo	CauEMO
Time	51.6	52.6	54.2	62.3	58.7	53.1

Further, the scalability can be also optimized with following aspects. 1) Divide-and-conquer scheme on graphs for prediction. For a larger molecule, we can extract the subgraphs and sub-blocks on such graph and impose the parallel computation to speed up the process. 2) Simplify the complexity of current CauEMO. Since our CauEMO enjoys the comparable and even better efficiency than other baselines, we can still provide further simplification by reducing the learnable parameters of stochastic attention in our CauEMO.

4.9 Interpretability discussion

In this subsection, we finally systematically evaluate the interpretability of our CauEMO from both theoretical and empirical perspectives.

- *Theoretical interpretability.* The design of CauEMO is inherently grounded in causal interpretability. The E-GIB module explicitly enforces a trade-off between preserving environment-relevant information and minimizing spurious correlations with the labels, which directly corresponds to identifying causal environment factors. The SCI (Soft Causal Interaction) module further introduces cross-attention combined with a learnable gating mechanism. This architecture ensures that the contribution of environment and invariant representations is adaptively re-weighted. When conflicting signals arise, the gating function allows down-weighting of misleading environmental cues, thus maintaining causal consistency.
- *Empirical interpretability.* We enhanced the interpretability analysis with case study visualizations. For example, in the synthetic conflict molecules, where the benzene ring determines the true label but the hydroxyl group spuriously suggests the opposite, atten-

tion heatmaps clearly show that CauEMO prioritizes the benzene substructure while suppressing the misleading OH group. Also, we have conducted knowledge-dropout experiments, where predefined functional groups were partially removed. Visualizations of the extracted subgraphs indicate that CauEMO still recovers chemically meaningful motifs, demonstrating robustness and interpretability even under incomplete prior knowledge from empirical aspect.

5 Conclusion

In this work, we propose a novel graph learning framework, CauEMO, to address the OOD challenges in molecule science, from the perspective of environment expansion and environment-invariance interactions on graphs. We systematically address such OOD prediction over molecular graph property on three aspects. First, to extend the scale of environment and ensure the augmented quality of molecules, a chemistry bond principle-based domain knowledge enhanced molecular environment generator is proposed. Second, for maximumly squashing the irrelevant information from the whole graph, we devise an Environment-GIB-based irrelevant environment disentanglement via deriving a modified environment-based graph information bottleneck, which not only decouples the causal invariant substructures, but also provides the interpretability and theoretical guarantee for our solution. Third, in order to allow sufficient information interactions between extracted environment and causal invariances, we further devise an environment-invariance soft causal interaction, which consists of a cross-attention mechanism for weighting the importance of environments and a gated causal bridge to enable dynamical interactions of two branches. We conduct extensive experiments on 6 datasets against 12 baselines including conventional graph learning backbones, subgraph-based backbones as well as environment-based learning backbones. The results on performance comparison and ablation studies demonstrate the overall superiority of our CauEMO and the effectiveness of each module in CauEMO. We believe our CauEMO can be a real interdisciplinary solution that intersects data mining, bio-informatics and chemistry with informative scientific insights.

In future, our research plans can be threefold. First, we will further investigate how to discover more causal invariant subgraphs to facilitate the OOD learning, i.e., more than one substructure in the whole graph, regarding properties from diverse chemistry and bioinformatic insights. Second, we are going to study how to design the fusion mechanism among multiple substructures as well as environments on graphs and enable a multi-substructure to multi-property research scheme. Third, the efficiency is still the permanent issue in learning-based mechanism, how to design and implement the divide-and-conquer mechanism to allow our model to adaptively adapt the newly arrived data, and extremely large molecules can be the following research direction for large-scale molecule datasets.

Acknowledgements This paper is partially supported by the National Natural Science Foundation of China (No.12227901, No.62502488) and Natural Science Foundation of Jiangsu Province (BK.20240460).

Author contribution Limin Li created the idea and wrote the manuscript, Kuo Yang implemented the experiment and model evaluation, Wenjie Du, Pengkun Wang help polish the paper, Yang Wang and Zhengyang Zhou supervise the research and provide the financial support.

Data availability The experimental data of this paper are open-sourced data available online, or from recent literature, including MUTAG [47], Open Graph Benchmark (OGB) [46] and DrugOOD benchmark [45]. The synthesized data are generated from SpuriousMotif. Thus, there is no ethical issue for such available data where no animals or human are involved.

Declarations

Conflict of interest The authors declare that there are no Conflict of interest for this research paper.

References

1. Shen X, Wang Y, Zhou K, Pan S, Wang X (2024) Optimizing ood detection in molecular graphs: A novel approach with diffusion models. In: proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining, pp. 2640–2650
2. Barbatti M, Ruckenbauer M, Plasser F, Pittner J, Granucci G, Persico M, Lischka H (2014) Newton-x: a surface-hopping program for nonadiabatic molecular dynamics. Wiley Interdiscipl Revi Comput Mol Sci 4(1):26–33
3. Du W, Yang X, Wu D, Ma F, Zhang B, Bao C, Huo Y, Jiang J, Chen X, Wang Y (2023) Fusing 2d and 3d molecular graphs as unambiguous molecular descriptors for conformational and chiral stereoisomers. Brief Bioinform 24(1):560
4. Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED (2023) Scaling deep learning for materials discovery. Nature 624(7990):80–85
5. Boiko DA, MacKnight R, Kline B, Gomes G (2023) Autonomous chemical research with large language models. Nature 624(7992):570–578
6. Zhang S, Liu X, Qi Z, Yan X, Yang W (2025) Gi-graph: a generative invariant graph learning scheme towards out-of-distribution generalization. IEEE Trans Knowl Data Eng 37:5934
7. Wang K, Liang Y, Li X, Li G, Ghanem B, Zimmermann R, Yi H, Zhang Y, Wang Y et al (2023) Brave the wind and the waves: discovering robust and generalizable graph lottery tickets. IEEE Trans Pattern Anal Mach Intell 46(5):3388–3405
8. Yang K, Zhou Z, Sun W, Wang P, Wang X, Wang Y (2023) Extract and refine: Finding a support subgraph set for graph representation. In: proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining, pp. 2953–2964
9. Wu Y, Wang X, Zhang A, He X, Chua T-S (2022) Discovering invariant rationales for graph neural networks. In: international conference on learning representations
10. Zheng X, Liu Y, Pan S, Zhang M, Jin D, Yu PS (2022) Graph neural networks for graphs with heterophily: A survey. arXiv preprint arXiv:2202.07082
11. Yi H, Jia L, Ding J, Li H (2024) Achieving material diversity in wire arc additive manufacturing: leaping from alloys to composites via wire innovation. Int J Mach Tools Manuf 194:104103
12. Bernstein AS (2014) Biological diversity and public health. Annu Rev Public Health 35(1):153–167
13. Wang S, Yang X, Islam R, Chen H, Xu M, Li J, Cai Y (2024) Enhancing distribution and label consistency for graph out-of-distribution generalization. In: 2024 IEEE international conference on data mining (ICDM), pp. 875–880. IEEE
14. Zhao T, Liu G, Wang D, Yu W, Jiang M (2022) Learning from counterfactual links for link prediction. In: international conference on machine learning, pp. 26911–26926. PMLR
15. Bui T-C, Li W-S (2023) Toward interpretable graph neural networks via concept matching model. In: 2023 IEEE international conference on data mining (ICDM), pp. 950–955. IEEE
16. Chen Y, Bian Y, Zhou K, Xie B, Han B, Cheng J (2023) Does invariant graph learning via environment augmentation learn invariance?
17. Yuan H, Sun Q, Fu X, Zhang Z, Ji C, Peng H, Li J (2024) Environment-aware dynamic graph learning for out-of-distribution generalization. Adv Neural Inf Process Syst 36:49715–49747
18. Xia Y, Liang Y, Wen H, Liu X, Wang K, Zhou Z, Zimmermann R (2024) Deciphering spatio-temporal graph forecasting: a causal lens and treatment. Adv Neural Inf Process Syst 36:37068–37088
19. Wang K, Liang Y, Wang P, Wang X, Gu P, Fang J, Wang Y (2022) Searching lottery tickets in graph neural networks: a dual perspective. In: the eleventh international conference on learning representations
20. Sun Q, Li J, Peng H, Wu J, Fu X, Ji C, Philip SY (2022) Graph structure learning with variational information bottleneck. In: proceedings of the AAAI conference on artificial intelligence, vol. 36, pp. 4165–4174
21. Gori M, Monfardini G, Scarselli F (2005) A new model for learning in graph domains. In: Proceedings. 2005 IEEE international joint conference on neural networks, 2005., vol. 2, pp. 729–734. IEEE
22. Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907
23. Hamilton W, Ying Z, Leskovec J (2017) Inductive representation learning on large graphs. Advances in neural information processing systems **30**

24. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y (2017) Graph attention networks. arXiv preprint arXiv:1710.10903
25. Alsentzer E, Finlayson S, Li M, Zitnik M (2020) Subgraph neural networks. Adv Neural Inf Process Syst 33:8017–8029
26. Sun Q, Li J, Peng H, Wu J, Ning Y, Yu PS, He L (2021) Sugar: Subgraph neural network with reinforcement pooling and self-supervised mutual information mechanism. In: proceedings of the web conference 2021, pp. 2081–2091
27. Yuan H, Tang J, Hu X, Ji S (2020) Xgnn: Towards model-level explanations of graph neural networks. In: proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 430–438
28. Lu Y, Zhang Y, Han B, Cheung Y-m, Wang H (2023) Label-noise learning with intrinsically long-tailed data. In: proceedings of the IEEE/CVF international conference on computer vision, pp. 1369–1378
29. Lu Y, Cheung Y-M, Tang YY (2019) Adaptive chunk-based dynamic weighted majority for imbalanced data streams with concept drift. IEEE Trans Neural Netw Learn Syst 31(8):2764–2778
30. Arjovsky M, Bottou L, Gulrajani I, Lopez-Paz D (2019) Invariant risk minimization. arXiv preprint arXiv:1907.02893
31. Li H, Wang X, Zhang Z, Zhu W (2022) Ood-gnn: out-of-distribution generalized graph neural network. IEEE Trans Knowl Data Eng 35(7):7328–7340
32. Du Y, Wang J, Feng W, Pan S, Qin T, Xu R, Wang C (2021) Adarnn: Adaptive learning and forecasting of time series. In: proceedings of the 30th ACM international conference on information & knowledge management, pp. 402–411
33. Zhou Z, Huang Q, Yang K, Wang K, Wang X, Zhang Y, Liang Y, Wang Y (2023) Maintaining the status quo: Capturing invariant relations for ood spatiotemporal learning. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 3603–3614
34. Pearl J (2009) Causal inference in statistics: An overview
35. Liu Y, Ao X, Feng F, Ma Y, Li K, Chua T-S, He Q (2023) Flood: A flexible invariant learning framework for out-of-distribution generalization on graphs. In: proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining, pp. 1548–1558
36. Yang N, Zeng K, Wu Q, Jia X, Yan J (2022) Learning substructure invariance for out-of-distribution molecular representations. Adv Neural Inf Process Syst 35:12964–12978
37. Wu Q, Zhang H, Yan J, Wipf D (2022) Handling distribution shifts on graphs: An invariance perspective. arXiv preprint arXiv:2202.02466
38. Wang Q, Guo B, Cheng L, Yu Z (2023) Surban: stable prediction for unseen urban data from location-based sensors. Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies 7(3):1–20
39. Wang B, Wang P, Xu W, Wang X, Zhang Y, Wang K, Wang Y (2024) Kill two birds with one stone: Rethinking data augmentation for deep long-tailed learning. In: the twelfth international conference on learning representations
40. Wu T, Ren H, Li P, Leskovec J (2020) Graph information bottleneck. Adv Neural Inf Process Syst 33:20437–20448
41. Yu J, Cao J, He R (2022) Improving subgraph recognition with variational graph information bottleneck. In: proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 19396–19405
42. Yu J, Xu T, Rong Y, Bian Y, Huang J, He R (2020) Graph information bottleneck for subgraph recognition. arXiv preprint arXiv:2010.05563
43. Alemi AA, Fischer I, Dillon JV, Murphy K (2016) Deep variational information bottleneck. arXiv preprint arXiv:1612.00410
44. Jang E, Gu S, Poole B (2016) Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144
45. Ji Y, Zhang L, Wu J, Wu B, Huang L-K, Xu T, Rong Y, Li L, Ren J, Xue D, et al (2022) Drugood: Out-of-distribution (ood) dataset curator and benchmark for ai-aided drug discovery–a focus on affinity prediction problems with noise annotations. arXiv preprint arXiv:2201.09637
46. Hu W, Fey M, Zitnik M, Dong Y, Ren H, Liu B, Catasta M, Leskovec J (2020) Open graph benchmark: datasets for machine learning on graphs. Adv Neural Inf Process Syst 33:22118–22133
47. Debnath AK, Compadre RL, Debnath G, Shusterman AJ, Hansch C (1991) Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds correlation with molecular orbital energies and hydrophobicity. J Med Chem 34(2):786–797
48. Xu K, Hu W, Leskovec J, Jegelka S (2018) How powerful are graph neural networks? arXiv preprint arXiv:1810.00826

49. Frasca F, Bevilacqua B, Bronstein M, Maron H (2022) Understanding and extending subgraph gnns by rethinking their symmetries. Adv Neural Inf Process Syst 35:31376–31390
50. Miao S, Liu M, Li P (2022) Interpretable and generalizable graph learning via stochastic attention mechanism. In: international conference on machine learning, pp. 15524–15543. PMLR
51. Chen Y, Zhang Y, Bian Y, Yang H, Kaili M, Xie B, Liu T, Han B, Cheng J (2022) Learning causally invariant representations for out-of-distribution generalization on graphs. Adv Neural Inf Process Syst 35:22131–22148
52. Jia T, Li H, Yang C, Tao T, Shi C (2024) Graph invariant learning with subgraph co-mixup for out-of-distribution generalization. Proceedings of the AAAI conference on artificial intelligence 38:8562–8570
53. Yang K, Zhou Z, Huang Q, Du W, Li L, Jiang W, Wang Y (2025) Enhancing graph invariant learning from a negative inference perspective. In: forty-second international conference on machine learning (ICML)
54. Yuan H, Sun Q, Fu X, Zhang Z, Ji C, Peng H, Li J (2023) Environment-aware dynamic graph learning for out-of-distribution generalization. Adv Neural Inf Process Syst 36:49715–49747