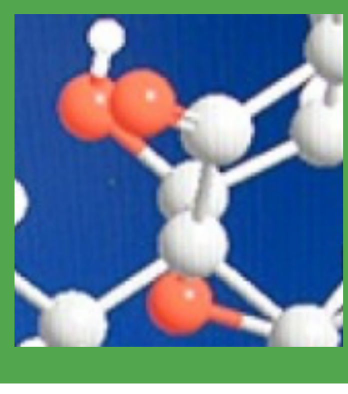
# Homological Representation Learning for Molecular Graphs



Yoshihiro Maruyama et al (Nagoya Uuniversity and Australian National University)

**Overview:** We propose Homological Rep. Learning (HomRL), an architecture-agnostic regularization for graph encoders that aligns latent embeddings with an efficiently computable homological signature of the input. On molecular-graph classification tasks, HomRL improves both in-distribution and OOD (out-of-distribution) accuracy.

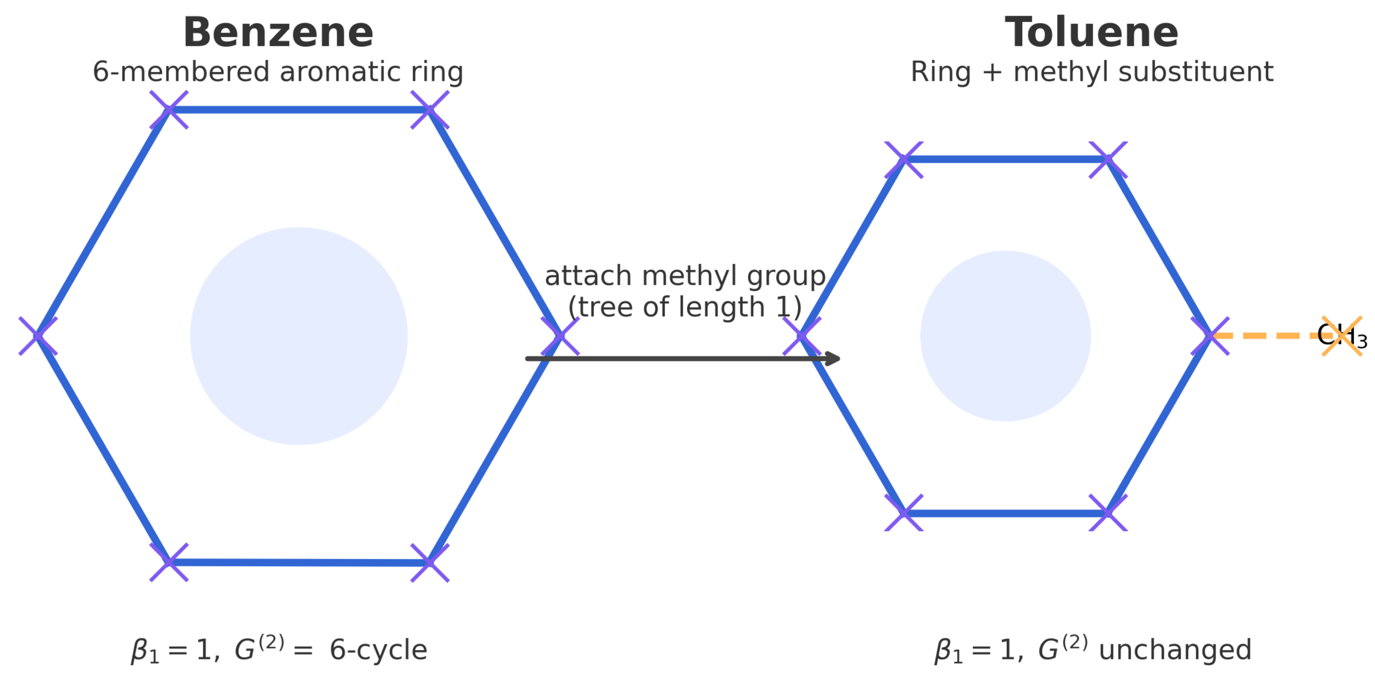
**Background:** A graph encoder produces an embedding $z_\theta(G) \in \mathbb{R}^d$ and a graph classifier predicts $\hat{y} = W_c z_\theta(G) + b_c$ . Many chemical structure edits preserve the molecular scaffold (e.g. ring structure). Standard loss functions do not provide stability under such edits.

**Homological Signature:**

$$s(G) \;=\; \Big(\beta_1(G),\ \chi(G),\ \frac{|V(G^{(2)})|}{|V(G)|},\ \#\mathrm{comp}(G^{(2)})\Big)$$

$$\widehat{s}(G) \;=\; W_m z_\theta(G) + b_m$$

Here we utilize the topological information such as Betti number, Euler characteristic, 2-core statistics for representation learning of chemical compounds




**Learning Objective:** We incorporate homological regularization terms into the objective function

$$\mathcal{L} \;=\; \mathrm{CE}(\hat{y}, y) \;+\; \lambda\,\|\widehat{s}(G) - s(G)\|_2^2 \;+\; \mu\,\mathbb{E}_{a\in\mathcal{A}}\big[\ \|\widehat{s}(a\cdot G) - \widehat{s}(G)\|_2^2\ \big]$$

**Experiment Results:**

| Model | Train Acc | Test Acc | OOD Acc |
|---|---|---|---|
| Baseline (no homology) | 0.6294 | 0.6517 | 0.6317 |
| HomRL (our method) | **0.6922** | **0.7000** | **0.6867** |

**Summary:** HomRL gives a regularization method that is easy to add to any graph encoder, comes with stability guarantees, and improves both in-distribution and OOD generalization at essentially no computational cost. The method can be combined with symmetry-aware architectures and richer algebraic topological information.

**Beyond:** This is the beginning of the story. We are applying categorical algebraic topology, such as categorical K-theory and motivic cohomology theory, for representation learning; algebraic topology is the source of interpretable representations for machine learning.

**Ref:** YM et al, Homological Representation Learning for Molecular Graphs, PMLR, 2025; YM et al, K-theoretic Persistent Cohomology, PMLR, 2025; YM, Algorithm for Interpretable Graph Features via Motivic Persistent Cohomology, Springer LNCS, 2025.