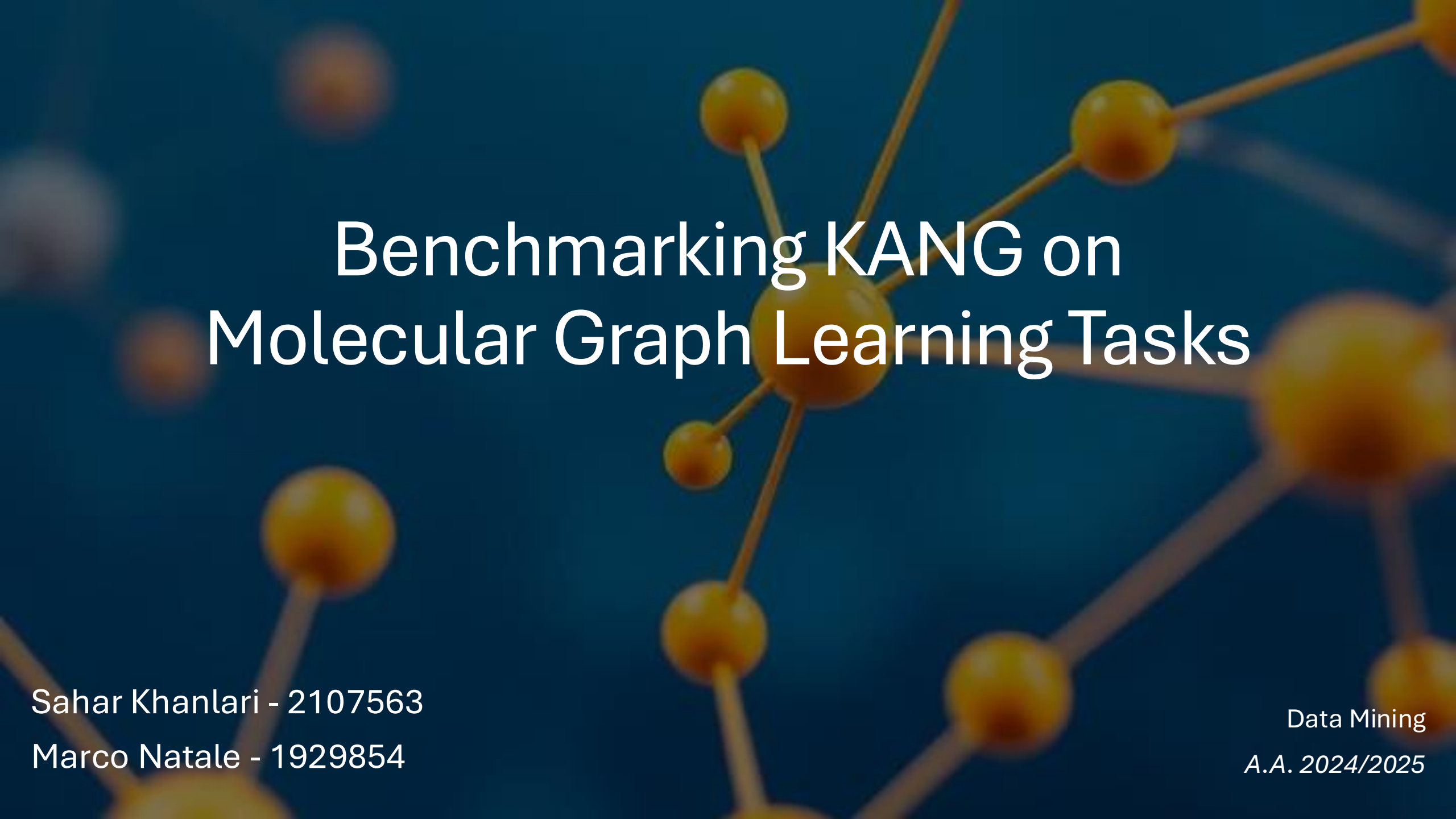


# Benchmarking KANG on Molecular Graph Learning Tasks

Sahar Khanlari - 2107563

Marco Natale - 1929854

Data Mining

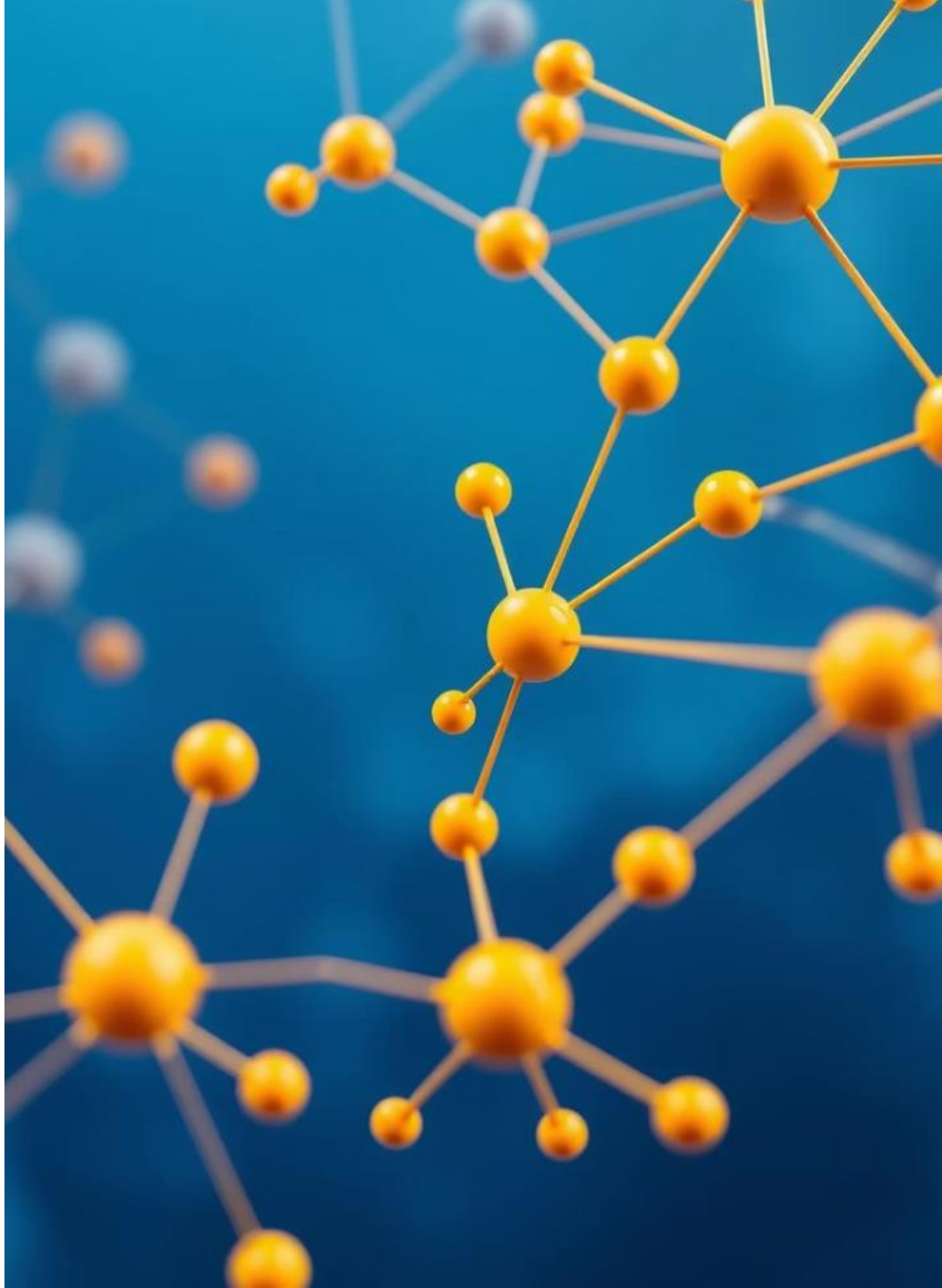
*A.A. 2024/2025*

# Introduction

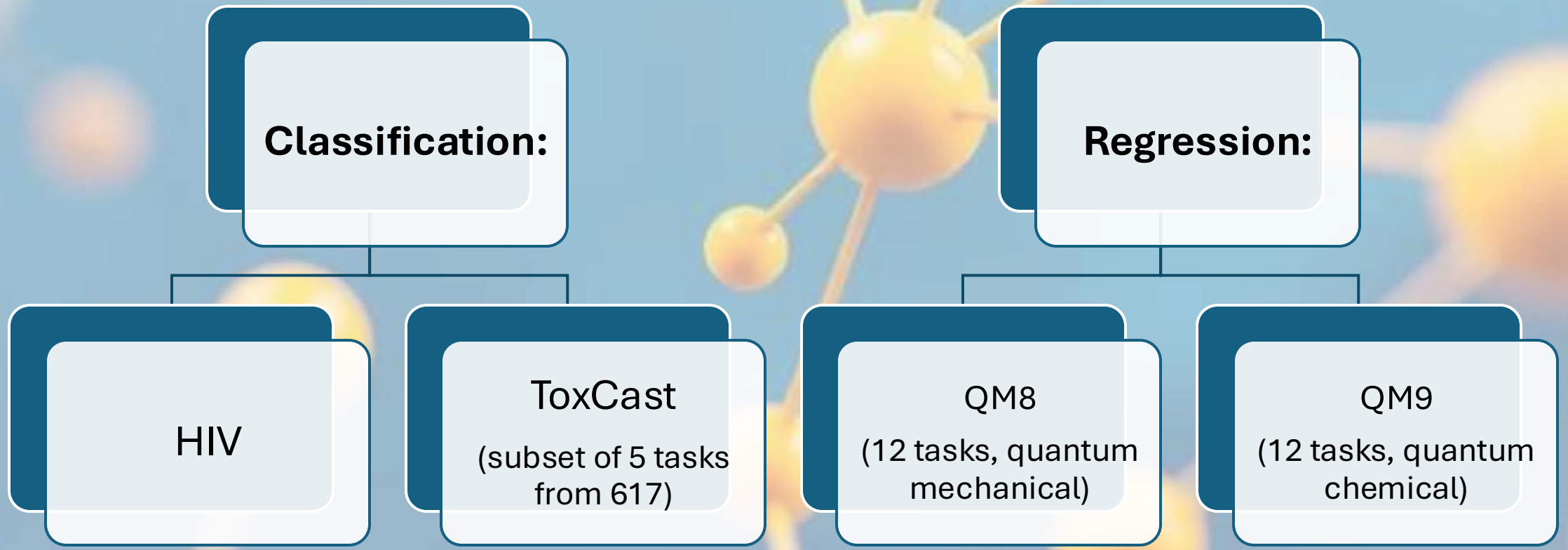
**Objective**: Evaluate the performance of the KANG model across classification and regression tasks using MoleculeNet datasets.

Molecular property prediction is crucial for drug discovery.

Only 2D molecular structures



# Overview of Datasets



# Experimental Setup

- SMILES to graph conversion with **RDKit**
- Graph generation: atoms → nodes, bonds → edges
- Hyperparameter tuning with **Optuna**
- Best val score saved and evaluated on test set

# Data Representation

**Node features:**

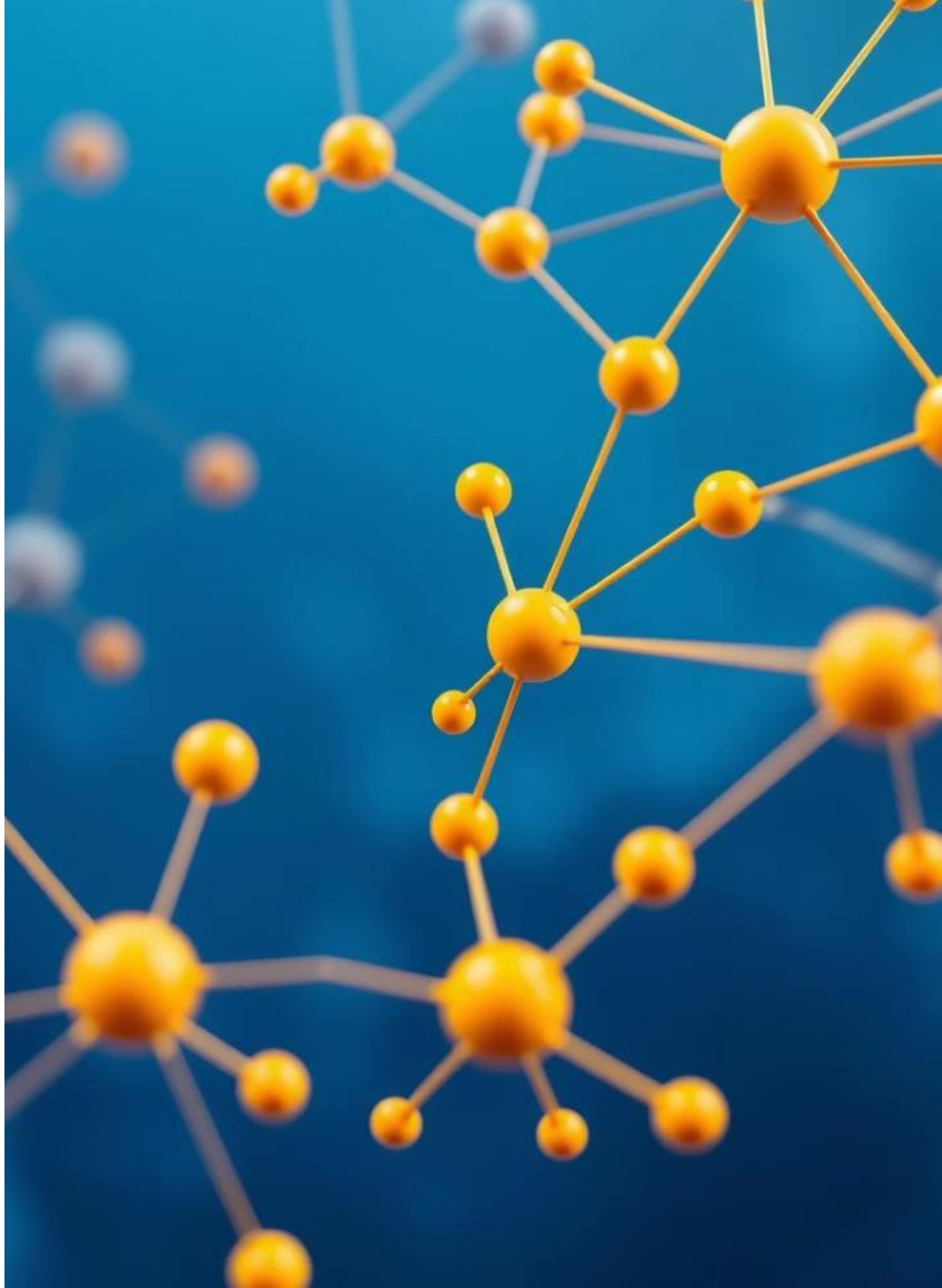
- Atomic number
- Degree
- Formal charge
- Aromaticity
- Total hydrogens

**Edge features:**

- Bond type
- Conjugation
- Ring membership
- Stereochemistry

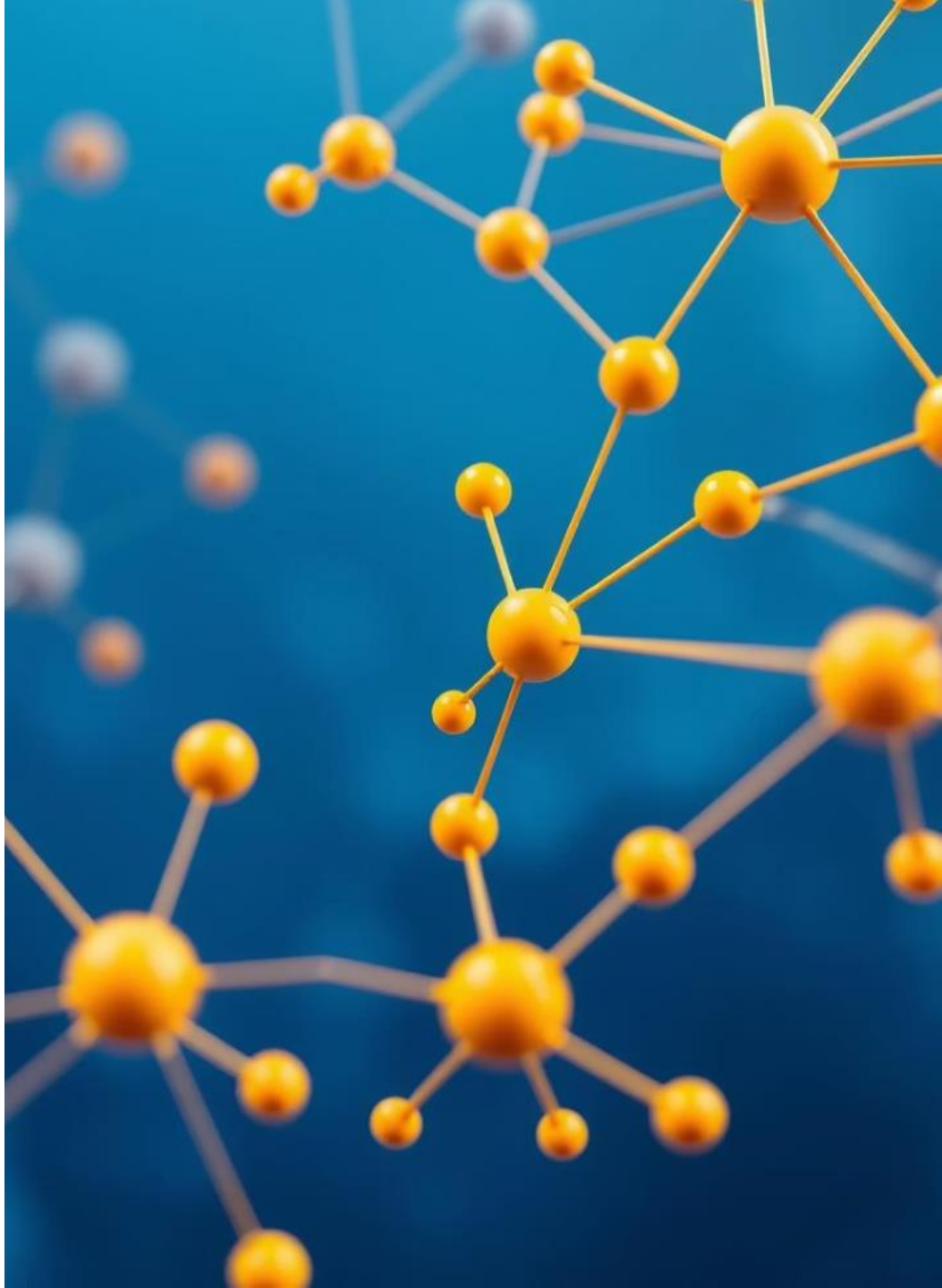
# Graph Classification - HIV

- Metric: ROC-AUC
- State-of-the-Art model: CIN++ (0.8063)
- **KANG result:** 0.6997



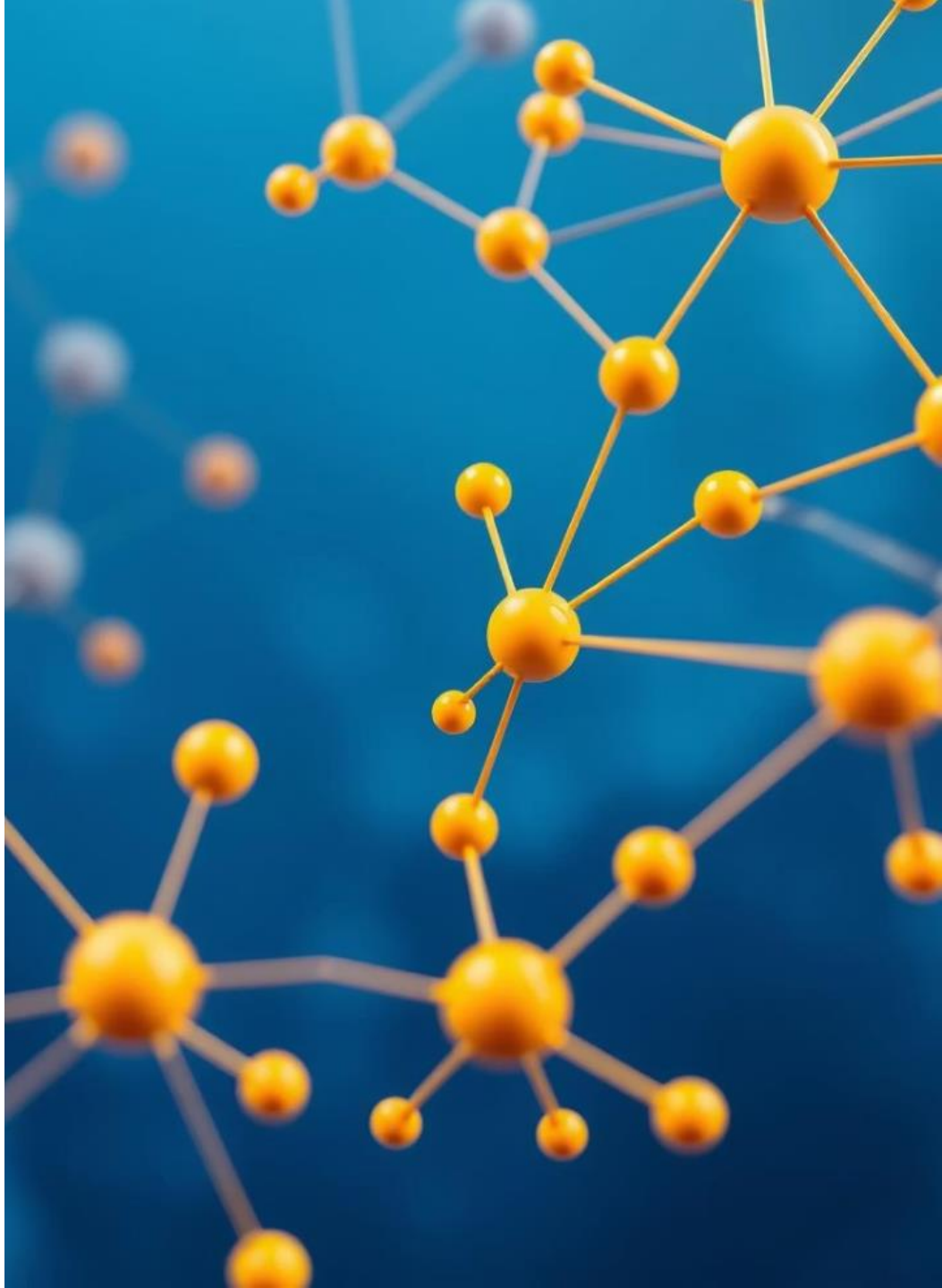
# Graph Classification - ToxCast

- 5 selected binary tasks from 617
- Tasks include: AhR, Aromatase, AutoFluor, p53 (2 types)
- State-of-the-Art model: DumplingGNN (0.782)
- **KANG average result:** 0.7896



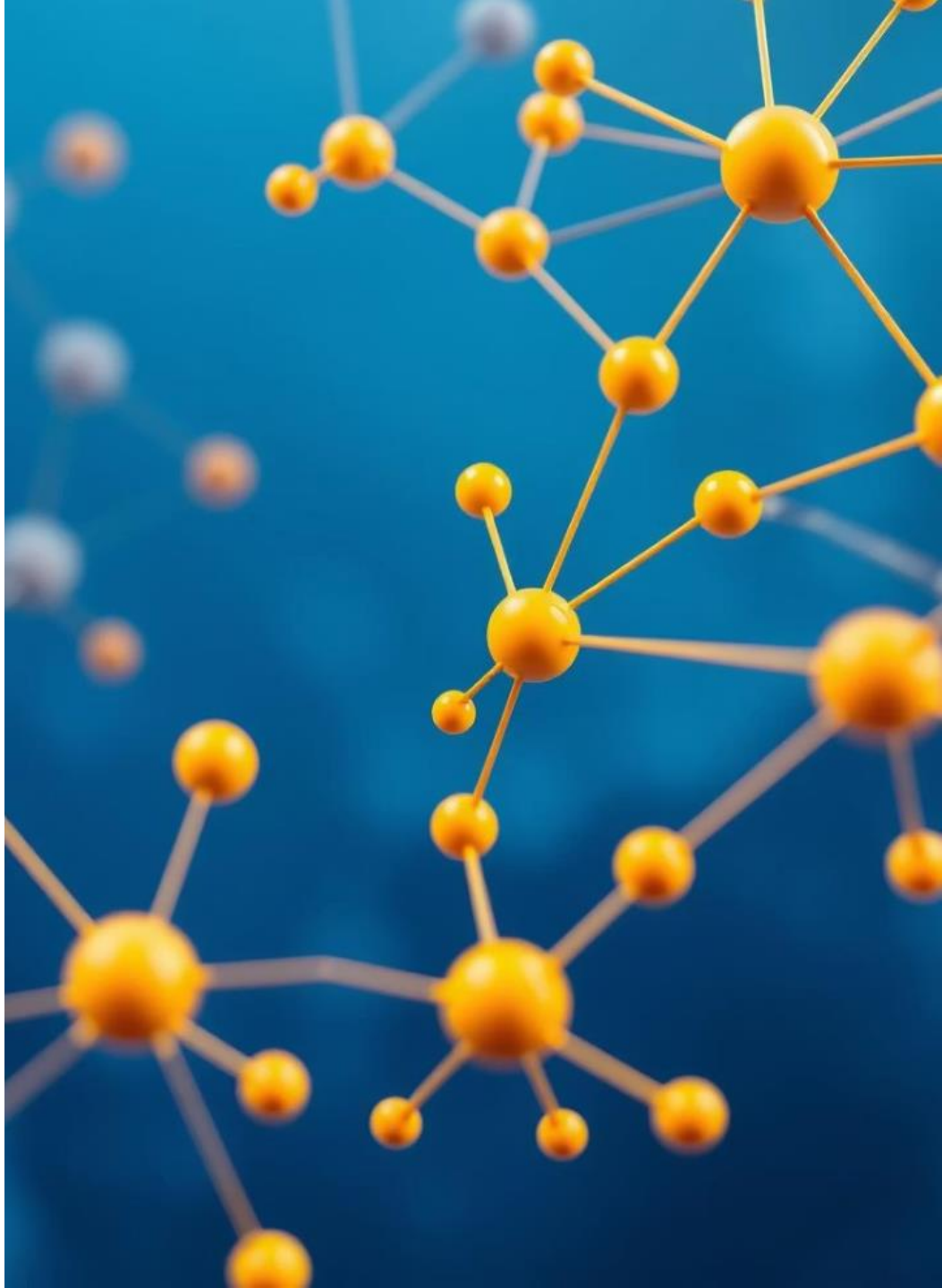
# Graph Regression - QM8

- 12 quantum mechanical tasks
- Metric: MAE
- SoTA: D-MPNN (0.0190)
- **KANG average result:** 0.0221



# Graph Regression - QM9

- 12 quantum chemical tasks
- Metric: MAE
- SoTA: D-MPNN (0.00814)
- **KANG average result:** 7.1407



# Performance Summary Table

| Dataset | Metric | SoTA Model | SoTA Value | KANG Value |
|---|---|---|---|---|
| QM8 | MAE | D-MPNN | **0.0190** | 0.0221 |
| QM9 | MAE | D-MPNN | **0.00814** | 7.1407 |
| HIV | ROC-AUC | CIN++ | **0.8063** | 0.6997 |
| ToxCast | ROC-AUC | DumplingGNN | 0.782 | 0.7896 |

# Conclusions

- Significant underperformance in QM9
- Comparable QM8 performance
- ToxCast comparable performance, but only 5 out of the total tasks were evaluated
- Slightly worse performance for HIV
- Further Improvements:
  - 3D Structure
  - Multi-Task Learning

# Thank You

Sahar Khanlari - 2107563

Marco Natale - 1929854

*A.A. 2024/2025*