Predicting Activity Coefficients at Infinite Dilution Using Hybrid Graph Neural Networks

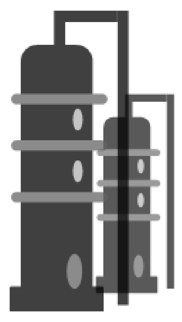
Edgar Ivan Sanchez Medina[1], Steffen Linke[1], Martin Stoll[2], Kai Sundmacher[1,3]

[1] Chair of Process Systems Engineering, Otto-von-Guericke University Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany
[2] Chair of Scientific Computing, Department of Mathematics, Technische Universität Chemnitz, 09107 Chemnitz, Germany
[3] Process Systems Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, 39106 Magdeburg, Germany

Motivation & Objective

Separation processes...

- consume 10-15% of the world's energy [1]
- constitute 40-50% of the total costs in chemical plants [2]



How to go beyond UNIFAC and COSMO-RS ?

We need accurate and efficient predictive thermodynamic methods!

Thermodynamics | Computational power | Data-driven tools

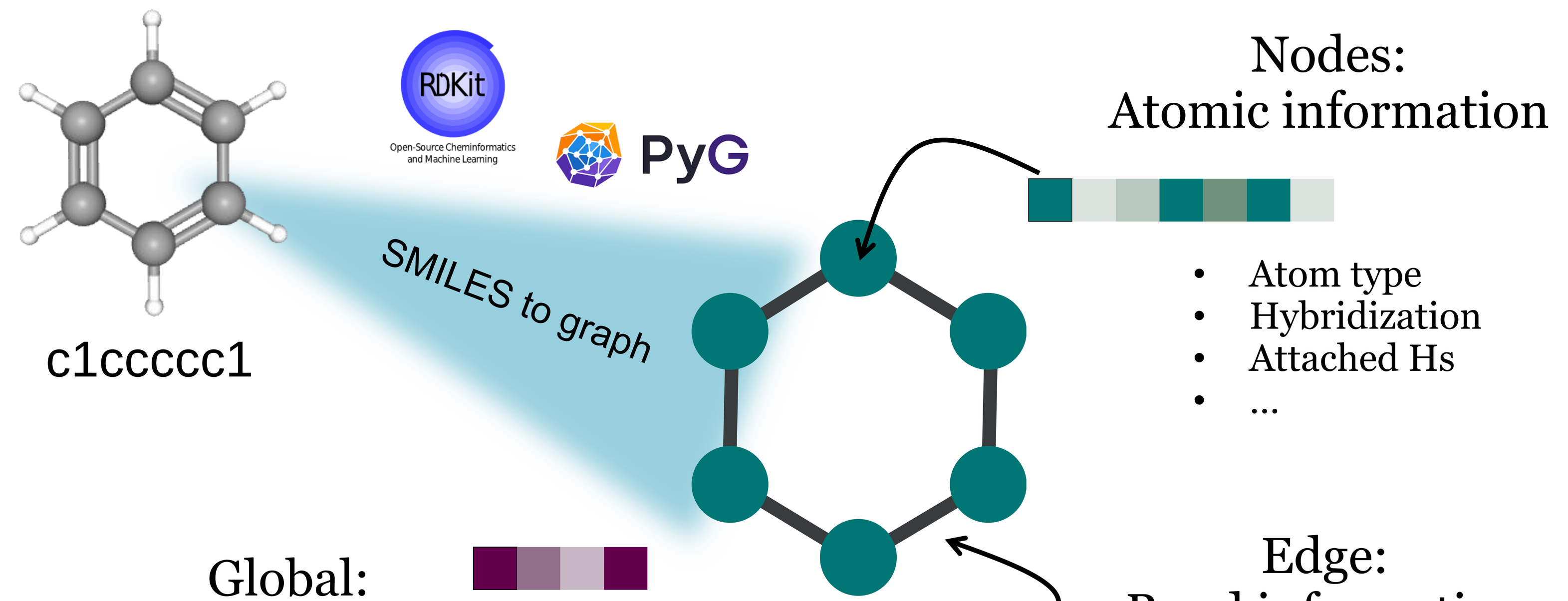
Can we developed such models now?

Activity coefficients at infinite dilution:

- High purity regimes (e.g., pollutants)
- Solvent selection for extractive distillation
- Activity coefficients at **finite** dilution

$$\gamma_{ij}^{\infty} = f\left(\text{[molecule]}, T \right)$$

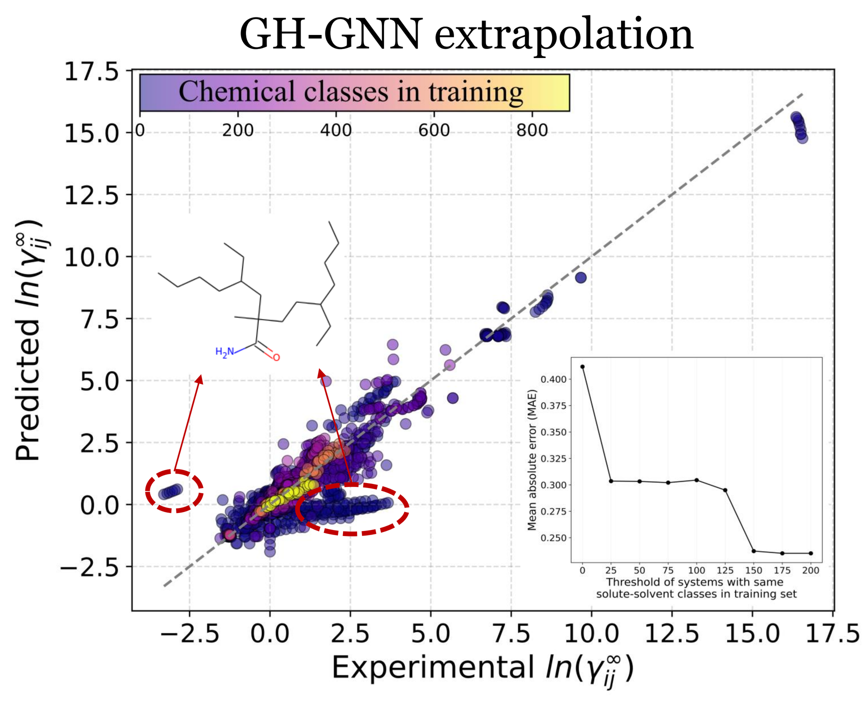
Molecular graphs



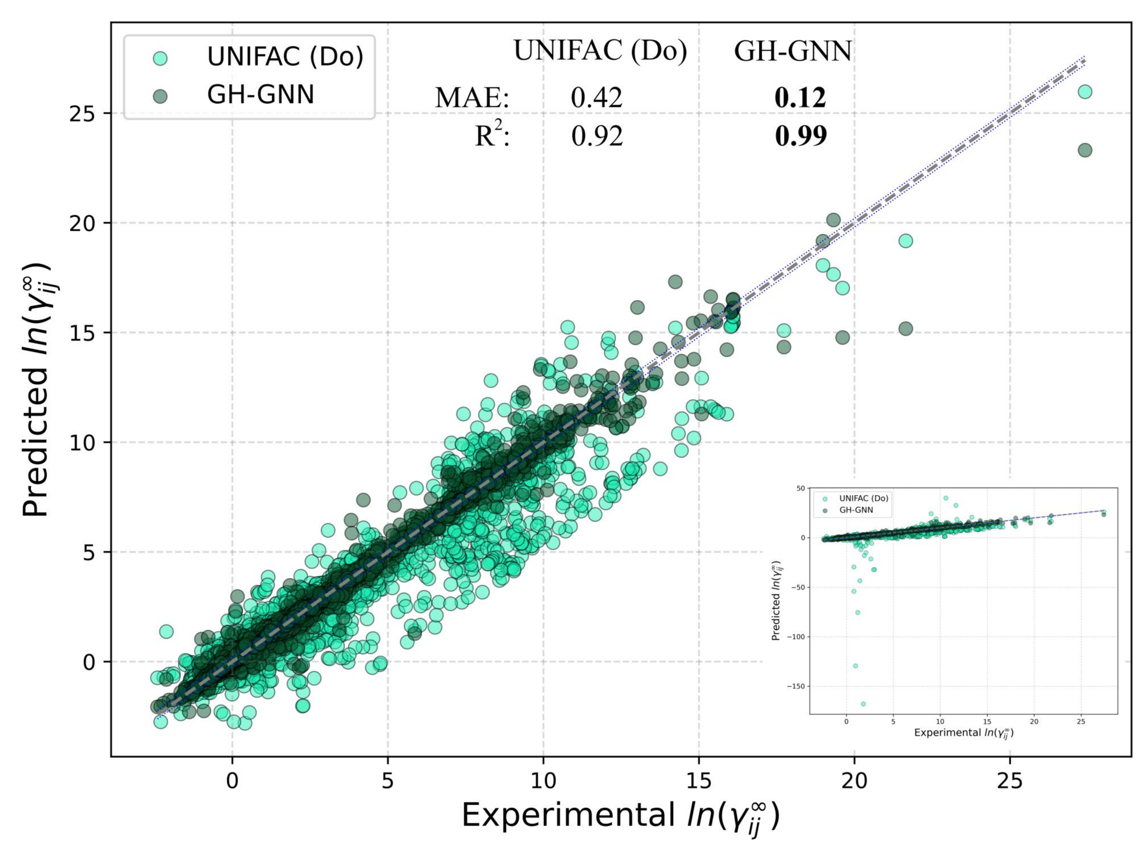
Hybrid serial Graph Neural Networks

DECHEMA experimental data

- 866 solvents and 1032 solutes
- 40,219 data points (1.64% of all binary systems)
- Stratified split (80/20) using chemical classes



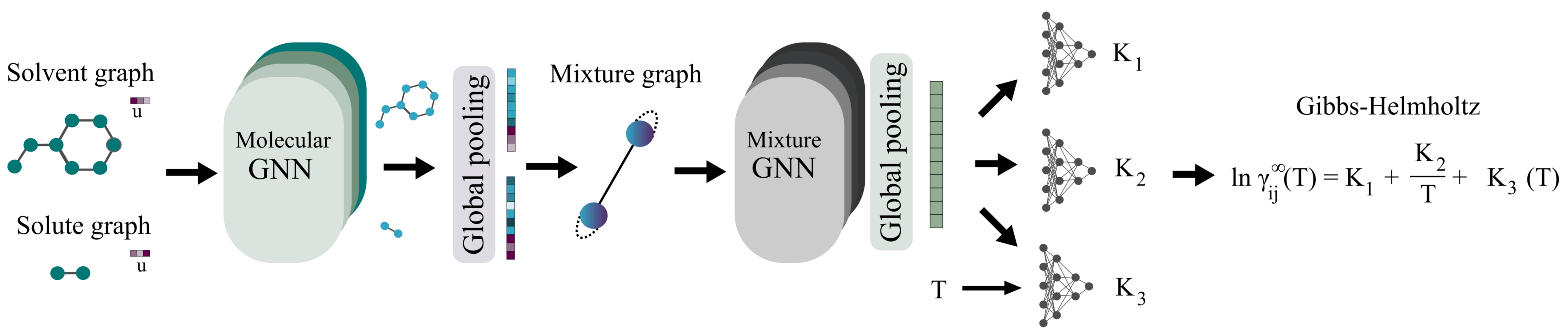
GNN ➡ Mechanistic ➡ $\ln \gamma_{ij}^{\infty}$



Gibbs-Helmholtz: Temperature dependency of γ_{ij}^{∞}

$$\frac{\partial\left(g_{ij}^{E}/RT\right)}{\partial T} = -\frac{h_{ij}^{E,\infty}}{RT^2} \rightarrow \frac{\partial\left(\ln \gamma_{ij}^{\infty}\right)}{\partial(1/T)} = -\frac{h_{ij}^{E,\infty}}{R}$$

Gibbs-Helmholtz Graph Neural Network (**GH-GNN**) [3]



- Metrics without the worse 1.6% UNIFAC (Do) predictions
- 16% of systems **cannot** be predicted with UNIFAC (Do)

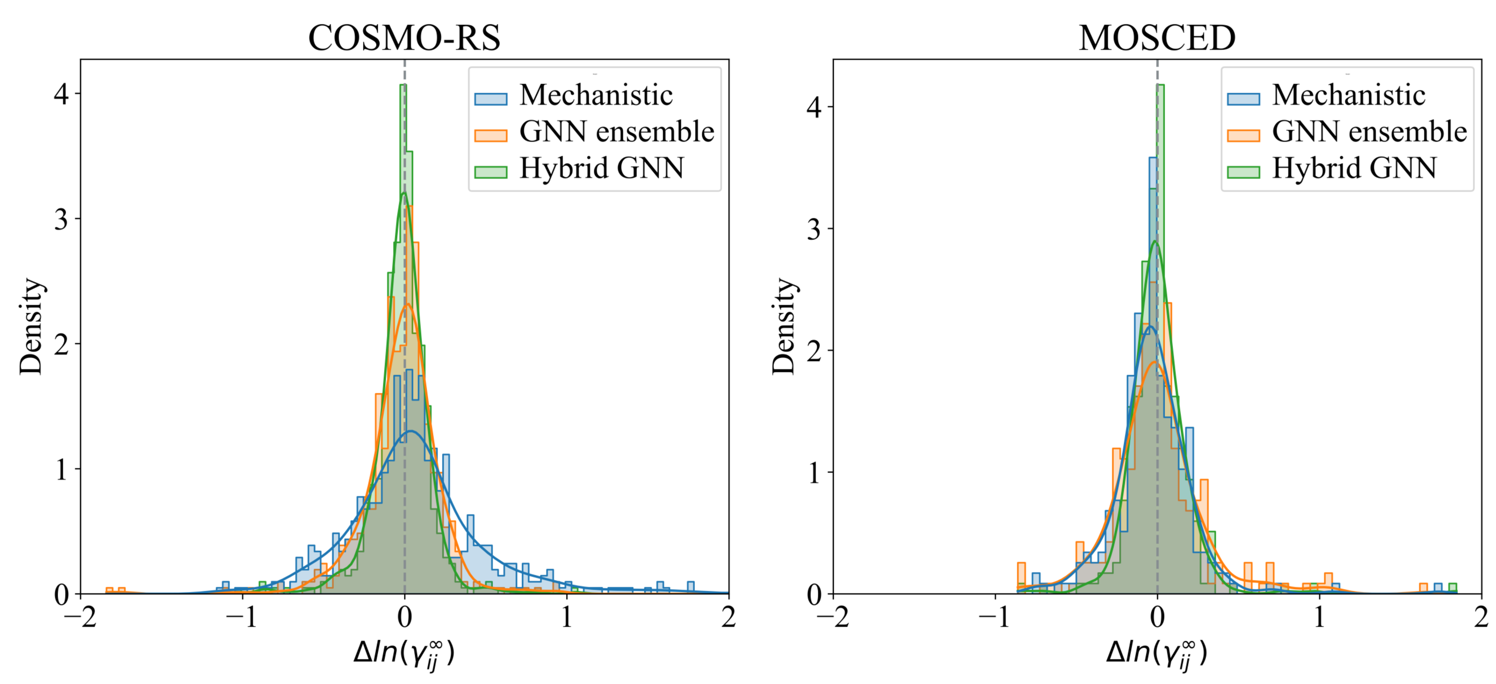
Hybrid parallel Graph Neural Networks

Isothermal GNN at 298.15 K [4]

- 262 solvents and 156 solutes
- 2810 data points (6.9% of all binary systems)
- Random split (80/20)
- Trained on solute and solvent final embedding concatenation
- Trained on the mechanistic models' residuals

$$r_{ij} = ln\left(\gamma_{ij}^{\infty,exp}\right) - \ln\left(\gamma_{ij}^{\infty,pred}\right)$$

- Ensemble learning to improve predictions



Mechanistic ➡ + ➡ $\ln \gamma_{ij}^{\infty}$; GNN ➡ +

Residual GNNs...

- can enhance prediction accuracy of mechanistic models
- are able to learn systematic errors in mechanistic models
- could be used for applicability domain prediction of well known models

Future work

- Serial hybrid GNN models to predict **phase-equilibria** (VLE, LLE, SLE)
- Methods for the **explainability** of mechanistic models' **residuals**
- Methods for applicability domain determination based on **multi-component** systems **similarity**
- Integration of **process development** pipelines

References

[1] Sholl D.S. and Lively R.P., Seven chemical separations to change the world. Nature, 532(7600):435–437, 2016.
[2] Kiss A.A., Lange J.P., Schuur B., Brilman D.W.F., van der Ham A.G.J., and Kersten S.R.A., Separation technology making a difference in biorefineries. Biomass and Bioenergy, 95:296–309, 2016.
[3] Sanchez Medina E.I., Linke S., Stoll M. and Sundmacher K., Gibbs-Helmholtz Graph Neural Network: Capturing the temperature dependency of activity coefficients at infinite dilution. Submitted. 2022.
[4] Sanchez Medina E.I., Linke S., Stoll M., and Sundmacher K., Graph neural networks for the prediction of infinite dilution activity coefficients. Digital Discovery, 1:216–225, 2022